A Multi-Modal Benchmark for Long-Range Depth Evaluation in Adverse Weather Conditions

Stefanie Walz*[1,3], Andrea Ramazzina*[2,4], Dominik Scheuble*[2,5], Samuel Brucker[1], Alexander Zuber[2], Werner Ritter[2], Mario Bijelic[1,6], Felix Heide[1,6]

***Abstract*— Depth estimation is a cornerstone computer vision application that is critical for scene understanding and autonomous driving. In real-world scenarios, achieving reliable depth perception under adverse weather—e.g. in fog and rain—is crucial to ensure safety and system robustness. However, quantitatively evaluating the performances of depth estimation methods in these scenarios is challenging due to the difficulty of obtaining ground truth data. A promising approach is using weather chambers to simulate diverse weather conditions in a controlled environment. However, current datasets are limited in distance and lack a dense ground truth.**
To address this gap, we introduce a novel evaluation benchmark that extends depth evaluation up to 200 meters under clear, foggy, and rainy conditions. To this end, we employ a multi-modal sensor setup, including state-of-the-art stereo RGB, RCCB, Gated camera systems, and a long-range LiDAR sensor. Moreover, we record a digital twin of the test facility sampled at a millimeter scale using a high-end geodesic laser scanner. This comprehensive benchmark allows for the evaluation of different models and multiple sensing modalities in a more precise and accurate manner, as well as at far distances. Data and code will be released upon publication.

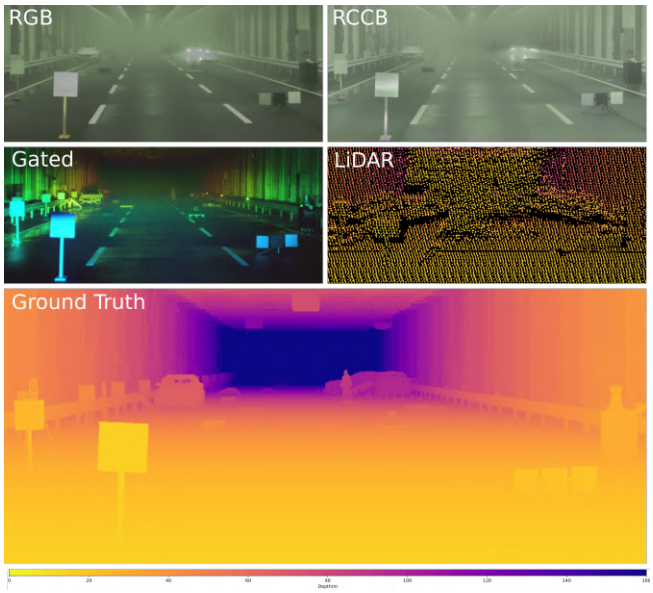


Fig. 1: We propose a multi-modal benchmark for depth evaluation in adverse weather conditions, covering ranges up to 200m. The benchmark includes state-of-the-art stereo RGB, RCCB, and Gated camera systems, as well as a long-range LiDAR sensor, each with pixel-wise annotated ground truth data.

I. Introduction

The development of safe autonomous driving systems depends on multi-modal sensor suites capable of providing comprehensive environmental perception through complementary sensing modalities. Contemporary solutions predominantly integrate camera, Light Detection and Ranging (LiDAR), and Radio Detection and Ranging (RaDAR) sensors, collectively enabling robust perception under nominal operating conditions. However, these sensor modalities face significant performance degradation in adverse weather scenarios due to interactions with airborne particulates such as precipitation or fog.

Under challenging meteorological conditions, camera-based systems suffer from diminishing image contrast caused by atmospheric light scattering [1], while LiDAR sensors experience spurious reflections from suspended water droplets that obscure critical object occlusion beyond precipitation zones [2]. In contrast, RaDAR maintains comparatively stable operation in such environments due to its superior penetration capabilities [3]. These limitations underscore the need for enhanced sensing paradigms that mitigate weather-induced artifacts without sacrificing perceptual accuracy.

Recent advances in automotive sensing address these challenges through innovative sensing modalities that exploit novel technological approaches and spectral regimes. Gated imaging systems, for instance, leverage a time-of-flight imaging principle to suppress backscatter by selectively capturing photons within temporally defined windows, demonstrating marked resilience to atmospheric obscurants [1].

However, thoroughly evaluating diverse sensor technologies necessitates appropriate testing methodologies and assessment criteria. While imaging sensors can be qualitatively assessed with relative ease, such evaluations often lack quantitative metrics to weigh advantages against limitations properly. Additionally, validating sensor models with real-world measurements presents difficulties due to the unpredictable and variable nature of adverse weather phenomena like fog or rain. Weather chambers [4], [5] have recently emerged as valuable testing environments, offering controlled and consistent fog and rain conditions for realistic sensor performance evaluation.

In this work, we focus on depth estimation and completion as essential elements for 3D scene understanding in autonomous driving safety systems. Depth information can be acquired through various approaches, namely passive methods, including stereo vision, structure from motion

* Denotes equal contribution, [1]Torc Robotics, [2]Mercedes-Benz AG, [3]Saarland University, [4]TU Munich, [5]TU Darmstadt, [6]Princeton University

(SfM), and monocular depth prediction, as well as active technologies like LiDAR and time-of-flight (ToF) cameras. However, adverse weather significantly degrades sensor performance, compromising the accuracy and reliability of depth estimation algorithms. Current datasets inadequately represent challenging weather conditions, limiting the ability to assess algorithm robustness in practical scenarios.

To address this gap, we introduce a depth evaluation framework featuring high-resolution annotated ground truth specifically designed for automotive applications in adverse weather. Our dataset is collected in a 200m-long fog chamber using multiple sensor modalities, as exemplified in Fig.1, ensuring reproducible and stable testing conditions. We benchmark leading algorithms across stereo depth estimation, monocular depth estimation, and sparse depth completion. Our analysis reveals not only the comparative strengths and weaknesses of different sensing approaches but also demonstrates how sensor degradation significantly impacts depth estimation accuracy. These findings contribute to developing more resilient autonomous driving systems capable of safe operation across diverse weather conditions.

Specifically, we make the following contributions:

- We introduce the first long-range (200m), multi-modal automotive depth evaluation dataset for adverse weather conditions with high-resolution ground truth data.
- We benchmark state-of-the-art depth estimation algorithms across multiple sensor modalities, including stereo matching, monocular depth prediction, and LiDAR-based depth completion.
- We analyze depth estimation performance under a wide range of adverse weather conditions, including fog, rain, and varying lighting, to highlight the strengths and pitfalls of different algorithms in realistic scenarios.

II. Related Work

A. Depth Estimation and Depth Completion

Computer vision encompasses a wide range of depth estimation approaches, each distinguished by its input modalities and measurement techniques. Depth estimation from intensity images includes both single-image [6], [7] and stereo methods [8]. Recent work has explored various approaches to enhance depth map resolution [9] and developed novel loss functions and neural network architectures [6], [7]. These methods typically rely on LiDAR data for training supervision [8] or employ self-supervised techniques to maintain stereo or temporal consistency [7], [6]. Depth completion research, meanwhile, utilizes sparse depth signals from sensors like LiDAR both for training supervision and as input data [10], [11], [12], [13], [14]. Some approaches combine LiDAR scans with synchronized RGB camera inputs to complete sparse depth point clouds [10], [11], [14], [13], while others incorporate stereo setups to enhance feature extraction and cost volumes [12]. Time-of-flight depth sensing technologies provide an alternative to conventional intensity images. These include correlation time-of-flight cameras [15], pulsed laser sensors [16], and gated illumination systems [17]. Pulsed time-of-flight sensors, such as LiDAR, measure light pulse travel time to and from the scene, offering high accuracy but limited spatial resolution and vulnerability to adverse weather [2], [18]. Gated cameras emit light pulses and capture reflections using a synchronized CMOS imager [19], [20], [21]. Recent studies combine deep learning with these systems to improve stereo depth estimation [20] and enhance accuracy through cross-spectral information [21].

B. Sensors in Adverse Environmental Conditions

Since adverse weather conditions degrade the performance of automotive sensors, recent research has focused on analyzing and mitigating their impact on perception systems. Hasirlioglu et al. [22] developed a theoretical model examining fog and rain's impact on camera, LiDAR, and radar systems. They validated their theory using specialized rain [23] and fog [24] simulators, demonstrating that LiDAR sensors are more susceptible to adverse weather than radar systems. Their findings also revealed decreased contrast in camera systems during adverse conditions. Multiple studies have attributed this contrast degradation to atmospheric particles scattering light in multiple directions [25], [26], [27], [28]. While various approaches attempt to enhance degraded images through physical models [26], polarization techniques [27], or normalized radiance calculations [28], their effectiveness often lacks objective measurement. Peynot et al. [29] compiled a dataset capturing challenging environmental conditions (dust, smoke, rain) using advanced autonomous vehicle perception systems. Their data revealed significant LiDAR performance degradation due to airborne dust particles detecting and obscuring obstacles behind dust clouds. Bijelic et al. [2] observed similar backscattering effects in foggy conditions. In contrast, 77 GHz automotive radar sensors show minimal susceptibility to backscatter effects [3]. Testing facilities like Cerema [4] and Carissma [22] enable sensor distortion characterization through reproducible adverse weather scenarios with adjustable intensity. These controlled environments allow systematic sensor performance evaluation but are limited by the facility constraints, most notably its dimensions. Additionally, initiatives like the Robust Vision Challenge encourage the development of resilient algorithms evaluated on diverse datasets, including synthetic extensions like Foggy Cityscapes [30]. However, real-world adverse weather evaluation remains challenging due to its unpredictable nature and rapid variations. To overcome the limitations of synthetic datasets, our work introduces a non-synthetic yet reproducible benchmark for depth estimation under controlled adverse weather conditions, including rain and fog, with a ground truth depth reaching 200 meters. This benchmark establishes a standardized framework for evaluating sensor performance in adverse weather, supporting the development of robust sensor systems for autonomous driving and related applications.

C. Adverse Weather Datasets

Adverse weather datasets are essential for advancing computer vision algorithms, particularly for scene understanding

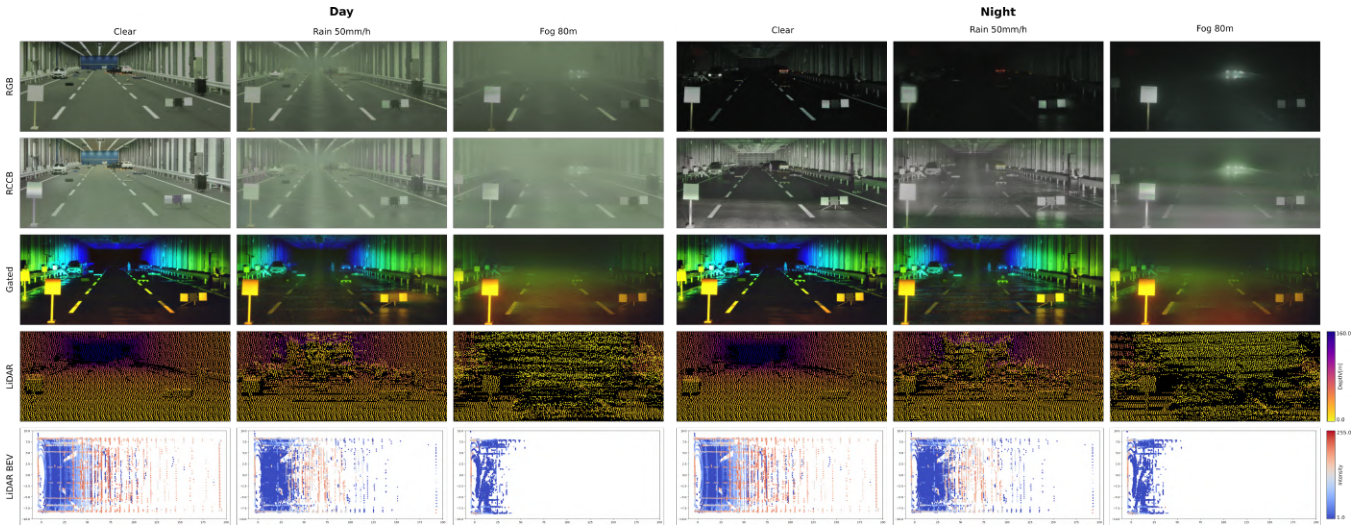


Fig. 2: We collected different scenarios in reproducible adverse weather conditions, during both day and night, in clear conditions, and with varying levels of fog and rainfall rates. We provide sensor measurements captured by three state-of-the-art camera systems: an RGB stereo system, an RCCB stereo system, and a Gated stereo system. Moreover, we also use a long-range LiDAR sensor.

and perception. Research efforts in this area have produced datasets simulating various weather conditions, including snowfall, rainfall [5], fog [30], and other environmental challenges. One common approach involves superimposing adverse weather effects onto clear weather images to create paired examples for supervised training. These simulation techniques often build upon Koschmieder's physical model [31]. However, synthesized datasets frequently lack the authenticity and variety found in real-world conditions. The RESIDE dataset [32] provides over 4000 real foggy scenes collected from the internet with object detection annotations, supporting research in dehazing and object detection under adverse conditions. The O-HAZE and I-HAZE datasets [33] feature real outdoor and indoor hazy scenes generated using professional haze machines, though their scope is limited by small sample sizes. Gruber et al. [34] proposed a depth benchmark dataset captured in a fog chamber under various adverse weather conditions, offering high-resolution ground truth depth. However, its 30-meter depth range limits its applicability to real-world scenarios. The STF dataset [35] encompasses 10,000 km of driving data across northern Europe in adverse weather conditions but relies solely on LiDAR for ground truth depth, which is compromised in adverse weather. Similarly, the LIBRE dataset [36] evaluates 10 LiDAR sensors in fog chamber conditions but provides only sparse LiDAR-based ground truth depth. In our work, we present a novel adverse weather dataset captured using a multi-modal sensor setup in a large-scale fog chamber, which provides high-resolution ground truth depth measurements extending to 200 meters.

III. Adverse Weather Dataset

This section introduces our adverse weather dataset captured in the Japan Automobile Research Institute (JARI) weather chamber [37]. We describe the chamber's capabilities, our methodology for ground truth depth data acquisition, and the scenarios investigated.

A. Weather Chamber Characteristics

The JARI weather chamber [37] is an advanced facility designed to simulate diverse weather conditions for research and testing. It extends 200 meters in length and 15 meters in width and features three 3.5-meter-wide lanes to accommodate various experimental configurations. The facility incorporates sophisticated lighting systems that provide adjustable horizontal illuminance from 200 to 1,600 lx, with dimming controls available at 50-meter intervals. The lighting system maintains a color temperature of 5,000K, creating daylight-white conditions ideal for precise testing. For rainfall simulation, the chamber employs dual sprinkler systems capable of producing rain droplets measuring 640 µm and 1400 µm. The system can generate three distinct rain intensities: strong (30 mm/h), intense (50 mm/h), and very intense (80 mm/h). When set to 50mm/h, the system can maintain continuous rainfall across the entire 200-meter length for 30 minutes. The rain area is split into two 100-meter sections for controlled experimentation. The chamber's fog simulation system utilizes 7.5 µm particles and offers adjustable visibility ranges from 10 to 80 meters. The system can maintain continuous fog emission across the whole 200-meter track for up to 60 minutes, ensuring consistent testing conditions.

B. Ground Truth Acquisition

We obtain high-precision ground truth depth measurements for static scenarios using a Leica ScanStation P30 laser scanner. This advanced system features a horizontal 360°/290° field of view, operates at 1550 nm wavelength, and can capture up to 1 million points per second with an angular accuracy of up to 8" and range accuracy of 1.2 mm + 10

TABLE I: Sensors specifications of the test vehicle.

Sensor	RGB Camera	RCCB Camera	Gated Camera	LiDAR
Make	OnSemi	OnSemi	BrightwayVision	Velodyne
Type	AR0230AT	AR0820AT	BrightEye	VLS128
Sensor Count	2	2	2	1
Quantization	12 bit	16 bit	10 bit	16 bit
Framerate	30 Hz	15 Hz	120 Hz	10 Hz
Resolution	1920×1024	3848×2176	1280×720	-
Field-of-View	39.6°×21.7°	52.8°×28.9°	31.1°×17.8°	360°×40°
Angular Resolution	0.020°	0.030°	0.024°	>0.1°
Wavelength	380-740 nm	380-1050 nm	808 nm	905 nm
Illumination	Passive	Passive	Active	Active

parts per million (ppm). To enhance resolution and minimize occlusions, we combine multiple point clouds captured from different overlapping positions. Each individual scan generates approximately 157M points and requires about 5 minutes to complete, which restricts this high-resolution acquisition method to a limited number of static scenes. All sensors in our automotive suite (shown in Fig. 3) are calibrated relative to this ground truth point cloud. We employ generalized iterative closest point (ICP) [38] to determine the rotation and translation transformation between the Leica laser scan and the Velodyne point cloud.

C. Adverse Weather Scenarios

For depth estimation algorithm evaluation, we construct realistic automotive scenarios incorporating cars, mannequins representing pedestrians, and lost cargo objects positioned at varying distances. We specifically model highway scenarios and construction sites to provide diverse and challenging evaluation environments. These scenes are captured under controlled weather and illumination conditions within the weather chamber. Daytime conditions are simulated with ceiling illumination levels of 200lx and 800lx, while nighttime conditions are created by turning off all lights. Notably, illumination does not affect the gated camera and LiDAR systems, as the artificial lighting operates in the visible spectrum rather than the NIR spectrum used by these sensors. After obtaining reference measurements in clear conditions, we fill the chamber with fog. We monitor fog density using meteorological visibility V, calculated as $V = -ln(0.05)/\beta$, where β represents atmospheric attenuation. To maintain consistent and uniform fog density throughout the 200m track, we implement a 30-minute waiting period before taking measurements. Data collection occurs under stable fog conditions at visibility levels of 20m, 40m, 60m, and 80m. Additional measurements are taken during fog dissipation at visibility levels of 100m, 150m, and 200m. For rainfall scenarios, we capture data at three intensity levels: light rain (30mm/h), strong rain (50mm/h), and intense rain (80mm/h). The complete test dataset comprises 20 samples per scenario (day/night) under clear conditions, plus samples for three rain intensities and seven fog visibility levels, totaling 660 test samples.

IV. Sensor Setup

To evaluate the accuracy of different depth prediction models, we built a test vehicle equipped with multiple sensor modalities. As shown in Fig. 3, the roof rack features two separate camera boxes for the perception system. The

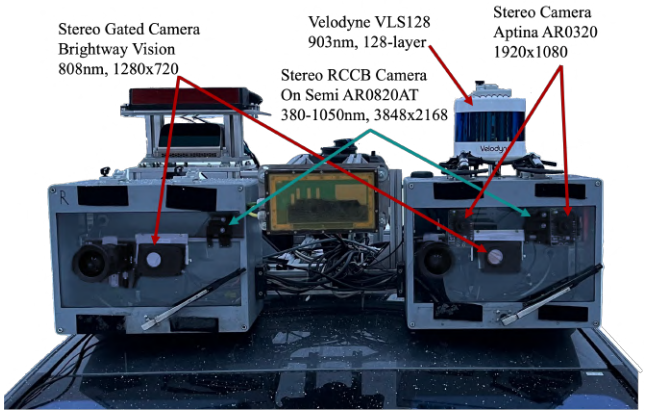


Fig. 3: Test vehicle equipped with a stereo RGB, RCCB, and Gated Camera, as well as a Velodyne VLS128 LiDAR.

left box houses a standard automotive stereo RGB camera system (OnSemi AR0230AT) with a 0.23 m narrow baseline. This system operates at 30 Hz, delivering 1920×1024 pixel resolution with 12 bit quantization and a field of view of 39.6°×21.7°. The left box also contains the left camera of a wide-baseline RCCB (OnSemi AR0820AT) and Gated (BrightWay Vision BrightEye) stereo camera system. The corresponding right cameras are mounted in the right box, enabling a 0.76 m baseline. Unlike conventional RGB cameras that use RGGB color filter arrays, RCCB cameras replace green channels with clear channels for enhanced night-time light sensitivity. The RCCB cameras operate at 15 Hz, with a 3848×2176 resolution and 16 bit quantization. The gated camera system differs as an active sensor, comprising two cameras and an illumination source mounted on the vehicle's towing hook. Its flood-light flash laser source illuminates the scene using variable light pulses lasting 240-370 ns at 808 nm wavelength. The system captures reflected light pulses using a synchronized camera after an adjustable delay, effectively minimizing particle backscatter in adverse weather to produce high-contrast images. The gated cameras feature a 31.1°×17.8° field of view and operate at 120 Hz. This high frame rate enables the capture of multiple overlapping slices with different range-intensity profiles, encoding implicit depth information. Following [20], we capture three active and two passive slices, achieving a 24 Hz final sampling rate. Gated images are captured at 1280×720 resolution with 10 bit quantization. The setup also includes a state-of-the-art Velodyne VLS128 long-range LiDAR system mounted above the left sensor box. Operating at 905 nm and 10 Hz, it provides a 40° vertical field of view with 128 non-linear distributed scanning lines at minimum 0.11° angular resolution. The system achieves ranges exceeding 200 m for highly reflective targets. Table I provides detailed technical specifications for all sensors.

V. Benchmark

In this section, we provide an overview of the state-of-the-art depth estimation methods used in our study and present the metrics used to quantitatively assess their performance.

TABLE II: Quantitative comparison of all benchmarked algorithms in clear and adverse weather conditions. Best results in each category are in **bold** and second best are underlined.

		clear					rain					fog				
	MODALITY [METHOD]	RMSE [m] ↓	MAE [m] ↓	δ_1 [%] ↑	δ_2 [%] ↑	δ_3 [%] ↑	RMSE [m] ↓	MAE [m] ↓	δ_1 [%] ↑	δ_2 [%] ↑	δ_3 [%] ↑	RMSE [m] ↓	MAE [m] ↓	δ_1 [%] ↑	δ_2 [%] ↑	δ_3 [%] ↑
DAY	RGB [39]	23.77	13.08	56.84	90.50	96.35	40.16	22.15	38.76	73.05	84.50	36.83	21.66	36.71	68.56	83.94
	Stereo-RGB [40]	6.81	5.11	98.43	99.53	99.82	14.27	8.29	87.15	95.31	96.99	29.03	16.02	71.86	80.80	86.14
	LiDAR-RGB [13]	6.84	2.75	97.53	99.30	99.75	30.33	11.27	85.15	89.55	91.56	55.62	36.05	32.02	36.72	41.33
	RCCB [39]	12.11	6.00	91.26	99.18	**99.89**	28.01	13.86	72.45	86.37	94.24	25.56	12.76	73.52	88.00	95.88
	Stereo-RCCB [40]	7.41	4.62	97.63	99.10	99.59	8.57	5.25	93.36	98.48	**99.61**	27.29	15.15	71.72	80.52	86.96
	LiDAR-RCCB [13]	7.20	2.97	96.82	98.97	99.76	31.98	12.27	82.69	87.54	90.20	56.78	37.47	29.90	34.63	39.56
	Stereo-Gated [20]	5.95	2.42	98.27	99.52	99.70	**8.31**	3.40	**97.46**	**99.37**	99.58	**11.30**	**5.41**	**92.73**	**98.28**	**99.43**
	Stereo-Gated-RCCB [21]	**5.22**	**1.68**	**98.84**	**99.58**	99.75	8.65	**3.20**	96.74	99.01	99.51	28.07	14.08	75.16	82.02	87.67
NIGHT	RGB [39]	23.72	14.18	47.84	86.85	96.99	36.42	20.48	28.50	73.22	85.27	42.47	26.33	11.19	55.40	76.48
	Stereo-RGB [40]	7.00	5.14	96.93	**99.58**	99.83	18.25	10.36	75.78	91.42	97.56	34.91	19.54	59.82	74.23	82.04
	LiDAR-RGB [13]	6.74	2.70	97.86	99.42	99.75	31.08	12.02	82.70	87.83	90.30	55.37	35.81	32.21	37.22	42.12
	RCCB [39]	14.21	7.12	88.15	98.82	**99.87**	32.02	16.37	65.01	83.27	91.82	25.93	12.60	77.04	90.17	95.66
	Stereo-RCCB [40]	7.64	4.74	97.10	99.12	99.69	9.56	5.28	94.20	98.31	99.47	33.46	19.55	58.45	69.40	78.04
	LiDAR-RCCB [13]	6.84	2.83	97.09	99.09	99.77	31.88	12.36	81.80	86.94	89.78	56.16	36.80	30.76	35.65	40.64
	Stereo-Gated [20]	6.06	2.49	98.36	99.51	99.70	**8.41**	3.63	**97.55**	**99.35**	**99.58**	**10.56**	**4.90**	**93.90**	**98.44**	**99.47**
	Stereo-Gated-RCCB [21]	**5.33**	**1.64**	**98.78**	99.56	99.73	9.58	**3.39**	96.04	98.98	99.50	29.84	14.12	75.99	82.79	88.26

A. Metrics

We conduct our evaluation using 2.5D depth images, as this format enables direct and intuitive comparison with ground truth data from depth cameras or projected LiDAR points. Our assessment employs established metrics from the KITTI benchmark [41], including Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), along with the threshold metric $\delta_i < 1.25^i$ for $i \in 1,2,3$.

B. Methods

While extensive research exists in depth estimation and completion, our benchmark focuses on one representative algorithm from each category for conciseness. In particular, we use DPT [39] as monocular RGB depth estimation model, CREStereo [40] as stereo approach, CompletionFormer [13] as LiDAR depth completion method, Gated Stereo [20] for (stereo) gated depth estimation and Gated-RCCB-Stereo [21] for cross-spectral stereo depth estimation. Additionally, DPT [42], CREStereo [40], and CompletionFormer [13] are trained and tested using both RGB and RCCB images as input to enable a comprehensive comparison between different camera modalities.

VI. ASSESSMENT

This section evaluates various depth estimation approaches using our benchmark dataset. We begin with an overview of our experimental methodology, followed by a comprehensive analysis of the methods' performance in both clear and adverse weather conditions.

A. Experimental Setup

Our evaluation framework follows the original training protocols proposed for each depth estimation method. We begin with the best available pre-trained models and further refine them using the Gated Stereo dataset [20], which employs an identical sensor configuration to our proposed adverse weather dataset. To address the absence of rain and fog scenarios in the existing dataset, we supplement the training data with additional samples captured in a fog chamber. Our augmented training set consists of 1080 samples that encompass diverse scenarios, including both clear and foggy conditions during day and night.

For our evaluation of the proposed test dataset, we compute metrics using pixel-wise annotated ground-truth depth for ranges between 0 and 200m. We evaluate all methods within the joint field of view of $23.81^\circ \times 10.38^\circ$ and a resolution of 980×420. We upsample all predicted depth maps using bilinear interpolation to match the predicted depth maps to the high-resolution ground truth.

B. Clear Weather Evaluation

We begin by analyzing the performance of all benchmarked algorithms under clear weather conditions. Fig.5 presents qualitative comparisons of predicted 2.5D depth maps across both daylight and nighttime scenarios. The results reveal significant limitations in RGB and RCCB monocular approaches for long-range depth estimation, with notable degradation in detecting distant objects such as vehicles and pedestrians at the far end of the fog chamber. In comparison, stereo and LiDAR depth completion methods demonstrate markedly improved accuracy at extended ranges. The Gated Stereo and Gated RCCB Stereo approaches show particularly robust performance during nighttime conditions, benefiting from the gated camera's active illumination system that enhances contrast and extends visible range. Table II presents our quantitative analysis. The RGB-based monocular method shows the lowest performance among all approaches. We observe that both RGB and RCCB-based monocular and stereo methods exhibit minor performance degradation during nighttime conditions, while LiDAR depth completion methods maintain consistent performance across lighting conditions. The Stereo Gated and Gated RCCB Stereo approaches achieve superior overall performance, successfully combining stereo and gated depth cues to produce accurate depth predictions with preserved fine details in both day and night.

C. Evaluation in Fog

In this section, we analyze the degradation patterns of depth estimation algorithms across increasing fog densities.

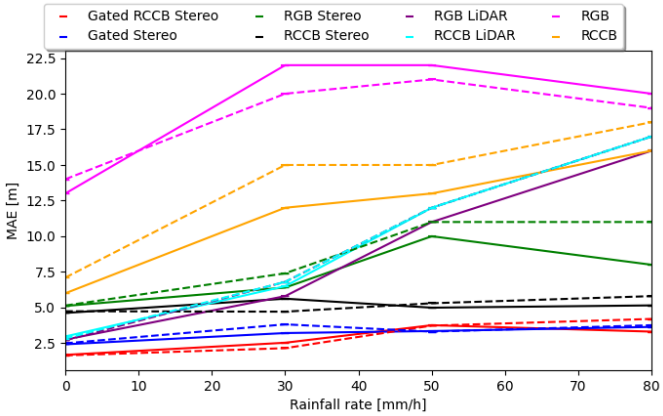
Our evaluation encompasses measurements from 7 distinct visibility levels, with qualitative results presented in Fig. 5. The quantitative analysis of mean-absolute error (MAE) across varying fog densities is illustrated in Fig. 4b. As fog density increases, camera images exhibit diminishing contrast, while LiDAR point clouds become increasingly contaminated with cluttered points due to severe back-scatter effects. Stereo-based approaches demonstrate relatively robust performance, showing only minimal degradation across most fog conditions, with significant performance drops occurring only at extremely low visibility levels. In contrast, LiDAR depth completion methods show high sensitivity to fog, with performance deteriorating even in light fog conditions due to the inability to differentiate between backscatter points from fog and valid measurements from actual objects. Gated Stereo emerges as the most robust solution across various fog conditions. The gating mechanism effectively mitigates backscatter in foggy conditions, enabling the detection of distant objects that become imperceptible in conventional RGB or RCCB streams, such as the pallet positioned at the fog chamber's far right. The Gated RCCB Stereo method maintains strong performance for visibility ranges exceeding 40m, though its effectiveness diminishes significantly at lower visibility ranges, primarily due to the degraded input quality from the supplementary RCCB camera.

D. Evaluation in Rain

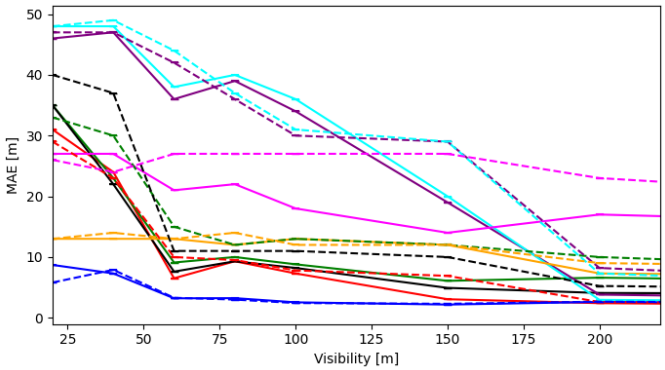
Our rain condition analysis encompasses three distinct rainfall intensities: 30mm/h, 50mm/h, and 80mm/h. Qualitative results are presented in Fig. 5, while Fig. 4a depicts the MAE across these varying rain intensities. Our findings indicate that rainfall generally has a less severe impact on depth estimation performance compared to foggy conditions. Monocular methods exhibit the highest sensitivity to rain, with notable performance degradation under heavy rainfall, particularly during nighttime operations. These methods are especially susceptible to errors caused by reflections from the road surface in wet conditions. Although LiDAR depth completion methods also show some performance deterioration in rain, the impact is considerably less pronounced than in monocular approaches. Stereo-based methods demonstrate remarkable resilience to increasing rainfall rates. While conventional RGB and RCCB stereo methods show some loss of detail at extended distances, both Gated Stereo and Gated RCCB Stereo maintain their ability to accurately estimate depth for distant objects, such as pedestrians and vehicles at the fog chamber's rear, demonstrating their robust performance in rainy conditions.

VII. Conclusion

We present the first comprehensive multi-modal depth evaluation dataset for long-range automotive applications, featuring high-resolution ground truth depth annotations with an angular resolution comparable to a 50MP camera. The dataset contains synchronized data from LiDAR and multiple stereo camera systems, capturing challenging highway scenarios under diverse weather conditions. Our benchmark



(a) Rain performance with respect to rainfall rate [mm/h].



(b) Fog performance with respect to visibility V [m].

Fig. 4: MAE with respect to visibility and rainfall rate, comparing performance during day (solid) and night (dashed).

enables a thorough comparative analysis of state-of-the-art depth estimation approaches, encompassing monocular and stereo methods, LiDAR depth completion techniques, and both gated stereo and cross-spectral stereo depth estimation algorithms. Our analysis reveals that stereo methods and LiDAR depth completion approaches achieve superior depth predictions for distant objects, while passive monocular depth estimation shows limitations due to its scale ambiguity from missing multi-view cues. In adverse weather evaluation, stereo-based approaches demonstrate significantly higher stability compared to LiDAR depth completion methods, which fail due to backscatter effects in fog and rain. Notably, gated camera-based methods maintain reliable performance even in nighttime and severe weather conditions, leveraging their active illumination and gating mechanisms. The weather chamber enables reproducible evaluation under adverse conditions, addressing a key limitation of existing datasets that often lack challenging weather scenarios due to sensor constraints. Our benchmark provides a unique platform for simultaneously assessing both the robustness and resolution of current depth estimation methods.

Future work could leverage the dataset to evaluate the generalization of simulation-trained models by testing them against real-world adverse weather conditions, helping identify which simulators yield the most transferable representations for safety-critical perception in autonomous driving.

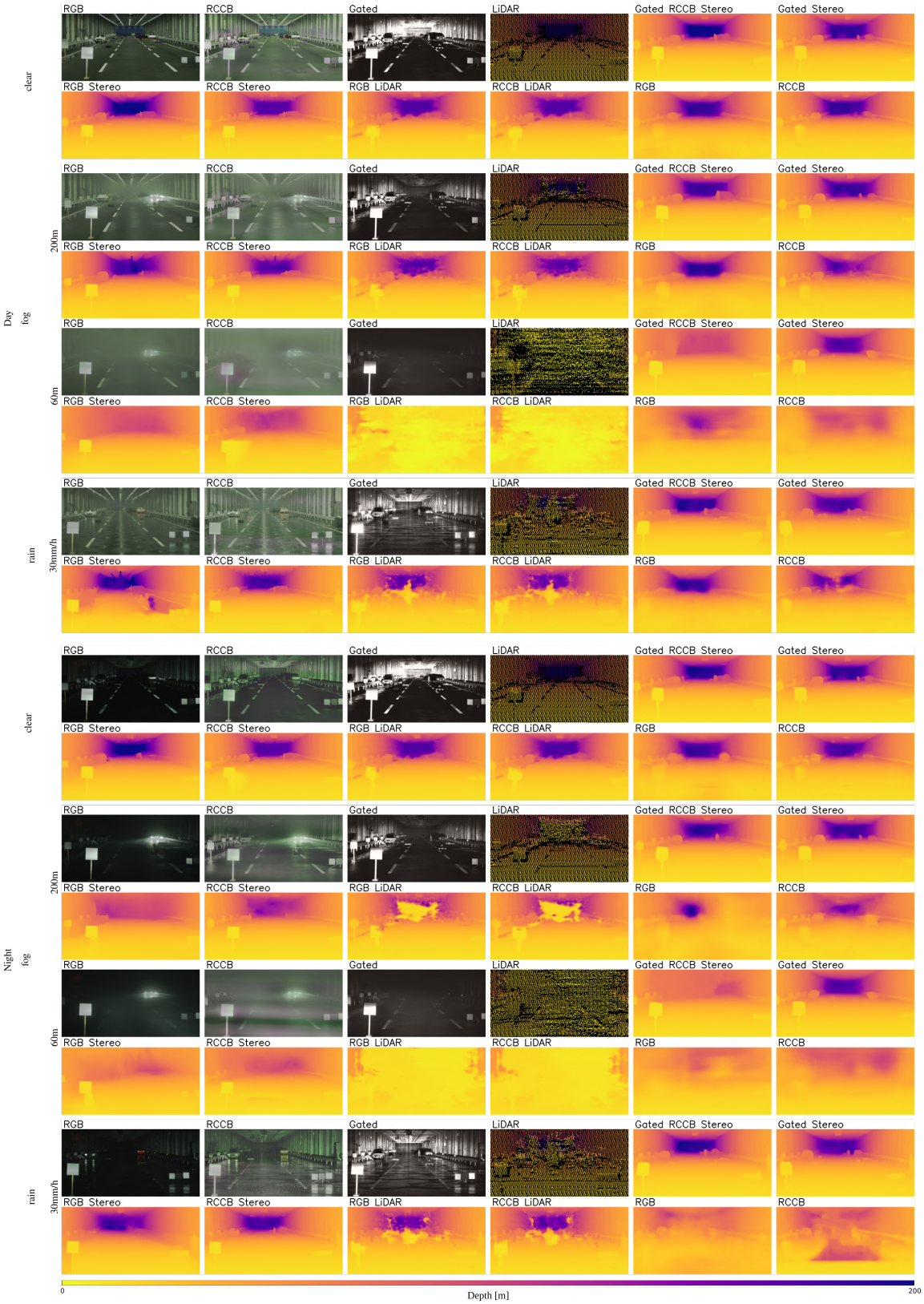


Fig. 5: Qualitative results for all benchmarked depth estimation methods in clear conditions, foggy conditions of different visibility levels and rainy conditions. In addition to the resulting 2.5D depth maps, we include a RGB, RCCB and Gated Image as well as the projected LiDAR point cloud as reference.

REFERENCES

[1] M. Bijelic, T. Gruber, and W. Ritter, "Benchmarking image sensors under adverse weather conditions for autonomous driving," in *IEEE Intelligent Vehicle Symposium*, 2018.

[2] G. T. Bijelic, Mario and W. Ritter, "A benchmark for lidar sensors in fog: Is detection breaking down?" in *2018 IEEE Intelligent Vehicles Symposium (IV)*, 2018, pp. 760–767.

[3] G. Brooker, R. Hennessey, C. Lobsey, M. Bishop, and E. Widzyk-Capehart, "Seeing through dust and water vapor: Millimeter wave radar sensors for mining applications," *Journal of Field Robotics*, vol. 24, no. 7, pp. 527–557, 2007.

[4] M. Colomb, J. Dufour, M. Hirech, P. Lacôte, P. Morange, and J.-J. Boreux, "Innovative artificial fog production device-a technical facility for research activities." in *The third international conference on fog, fog collection and dew*, 2004.

[5] S. Hasirlioglu and A. Riener, "A general approach for simulating rain effects on sensor data in real and virtual environments," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 426–438, 2019.

[6] C. Godard, O. Mac Aodha, and G. J. Brostow, "Unsupervised monocular depth estimation with left-right consistency," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 270–279.

[7] V. Guizilini, R. Ambrus, S. Pillai, A. Raventos, and A. Gaidon, "3d packing for self-supervised monocular depth estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2485–2494.

[8] J.-R. Chang and Y.-S. Chen, "Pyramid stereo matching network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5410–5418.

[9] S. M. H. Miangoleh, S. Dille, L. Mai, S. Paris, and Y. Aksoy, "Boosting monocular depth estimation models to high-resolution via content-adaptive multi-resolution merging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9685–9694.

[10] A. Wong and S. Soatto, "Unsupervised depth completion with calibrated backprojection layers," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 747–12 756.

[11] J. Park, K. Joo, Z. Hu, C.-K. Liu, and I. S. Kweon, "Non-local spatial propagation network for depth completion," in *Proc. of European Conference on Computer Vision (ECCV)*, 2020.

[12] J. Choe, K. Joo, T. Imtiaz, and I. S. Kweon, "Volumetric propagation network: Stereo-lidar fusion for long-range depth estimation," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 4672–4679, 2021.

[13] Y. Zhang, X. Guo, M. Poggi, Z. Zhu, G. Huang, and S. Mattoccia, "Completionformer: Depth completion with convolutions and vision transformers," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 18 527–18 536.

[14] T. Wu, B. Vallet, and M. Pierrot-Deseilligny, "Psmnet-fusionx3: Lidar-guided deep learning stereo dense matching on aerial images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 6526–6535.

[15] A. Kolb, E. Barth, R. Koch, and R. Larsen, "Time-of-flight cameras in computer graphics," in *Computer Graphics Forum*, vol. 29. Wiley Online Library, 2010, pp. 141–159.

[16] B. Schwarz, "Lidar: Mapping the world in 3D," *Nature Photonics*, vol. 4, no. 7, p. 429, 2010.

[17] P. Heckman and R. T. Hodgson, "Underwater optical range gating," *IEEE Journal of Quantum Electronics*, pp. 445–448, 1967.

[18] A. Carballo, J. Lambert, A. Monrroy, D. Wong, P. Narksri, Y. Kitsukawa, E. Takeuchi, S. Kato, and K. Takeda, "Libre: The multiple 3d lidar dataset," in *IEEE Intelligent Vehicles Symposium (IV)*, 2020.

[19] A. Walia, S. Walz, M. Bijelic, F. Mannan, F. Julca-Aguilar, M. Langer, W. Ritter, and F. Heide, "Gated2gated: Self-supervised depth estimation from gated images," 2022.

[20] S. Walz, M. Bijelic, A. Ramazzina, A. Walia, F. Mannan, and F. Heide, "Gated stereo: Joint depth estimation from gated and wide-baseline active stereo cues," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 252–13 262.

[21] S. Brucker, M. Bijelic, S. Walz, and F. Heide, "Cross-spectral gated-rgb stereo depth estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2024.

[22] S. Hasirlioglu and A. Riener, "Challenges in object detection under rainy weather conditions," in *Intelligent Transport Systems, From Research and Development to the Market Uptake: Second EAI International Conference, INTSYS 2018, Guimarães, Portugal, November 21–23, 2018, Proceedings 2*. Springer, 2019, pp. 53–65.

[23] S. Hasirlioglu, A. Kamann, I. Doric, and T. Brandmeier, "Test methodology for rain influence on automotive surround sensors," in *IEEE Int. Conf. on Intelligent Transportation Systems*. IEEE, 2016, pp. 2242–2247.

[24] S. Hasirlioglu, I. Doric, A. Kamann, and A. Riener, "Reproducible fog simulation for testing automotive surround sensors," in *2017 IEEE 85th Vehicular Technology Conference (VTC Spring)*. IEEE, 2017, pp. 1–7.

[25] M. Bijelic, T. Gruber, and W. Ritter, "Benchmarking image sensors under adverse weather conditions for autonomous driving," in *IEEE Intelligent Vehicle Symposium*. IEEE, 2018, pp. 1773–1779.

[26] J. P. Oakley and B. L. Satherley, "Improving image quality in poor visibility conditions using a physical model for contrast degradation," *IEEE transactions on image processing*, vol. 7, no. 2, pp. 167–179, 1998.

[27] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, "Instant dehazing of images using polarization," in *Conf. on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2001, pp. I–I.

[28] S. G. Narasimhan and S. K. Nayar, "Removing weather effects from monochrome images," in *Conf. on Computer Vision and Pattern Recognition*, vol. 2. IEEE, 2001, pp. II–II.

[29] T. Peynot, J. Underwood, and S. Scheding, "Towards reliable perception for unmanned ground vehicles in challenging conditions," in *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*. IEEE, 2009, pp. 1170–1176.

[30] C. Sakaridis, D. Dai, and L. Van Gool, "Semantic foggy scene understanding with synthetic data," *International Journal of Computer Vision*, vol. 126, pp. 973–992, 2018.

[31] H. Israël, F. Kasten, H. Israël, and F. Kasten, "Koschmieders theorie der horizontalen sichtweite," *Die Sichtweite im Nebel und die Möglichkeiten ihrer künstlichen Beeinflussung*, pp. 7–10, 1959.

[32] B. Li, W. Ren, D. Fu, D. Tao, D. Feng, W. Zeng, and Z. Wang, "Benchmarking single-image dehazing and beyond," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 492–505, 2018.

[33] C. Ancuti, C. O. Ancuti, R. Timofte, and C. De Vleeschouwer, "I-haze: a dehazing benchmark with real hazy and haze-free indoor images," in *Advanced Concepts for Intelligent Vision Systems: 19th International Conference, ACIVS 2018, Poitiers, France, September 24–27, 2018, Proceedings 19*. Springer, 2018, pp. 620–631.

[34] T. Gruber, M. Bijelic, F. Heide, W. Ritter, and K. Dietmayer, "Pixel-accurate depth evaluation in realistic driving scenarios," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2019, pp. 95–105.

[35] M. Bijelic, T. Gruber, F. Mannan, F. Kraus, W. Ritter, K. Dietmayer, and F. Heide, "Seeing through fog without seeing fog: Deep multimodal sensor fusion in unseen adverse weather," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 6 2020.

[36] A. Carballo, J. Lambert, A. Monrroy, D. Wong, P. Narksri, Y. Kitsukawa, E. Takeuchi, S. Kato, and K. Takeda, "Libre: The multiple 3d lidar dataset," in *2020 IEEE intelligent vehicles symposium (IV)*. IEEE, 2020, pp. 1094–1101.

[37] [Online]. Available: https://www.jari.or.jp/en/test-courses/jtown/46111/

[38] P. J. Besl and N. D. McKay, "Method for registration of 3-d shapes," in *Sensor fusion IV: control paradigms and data structures*, vol. 1611. Spie, 1992, pp. 586–606.

[39] R. Ranftl, A. Bochkovskiy, and V. Koltun, "Vision transformers for dense prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12 179–12 188.

[40] J. Li, P. Wang, P. Xiong, T. Cai, Z. Yan, L. Yang, J. Liu, H. Fan, and S. Liu, "Practical Stereo Matching via Cascaded Recurrent Network with Adaptive Correlation," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. New Orleans, LA, USA: IEEE, Jun 2022, pp. 16 242–16 251.

[41] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[42] R. Ranftl, V. Vineet, Q. Chen, and V. Koltun, "Dense monocular depth estimation in complex dynamic scenes," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 4058–4066.