AI-Generated Synthetic Data For Training Autonomous Vehicles In Extreme Conditions

Adarsh Velmurugan[1], Nivetha Elangovan[2]
[1, 2] Dept of Artificial Intelligence and Data Science
[1, 2] Chennai, Tamil Nadu, India

Abstract- *Advancements in autonomous vehicle technology rely heavily on their ability to navigate diverse and challenging conditions, including extreme environments. This paper explores the use of AI-generated synthetic data to enhance training datasets for autonomous vehicles, focusing on scenarios difficult to capture in real-world settings, such as severe weather, low-light conditions, and unusual road situations. Utilizing generative adversarial networks (GANs) and other AI techniques, synthetic data is created to closely mimic real-world conditions, enhancing the realism and diversity of training data. By integrating synthetic data with real-world datasets, autonomous systems can be trained more comprehensively, improving their robustness and safety by exposing them to a wider range of challenging scenarios during training. High-quality labels of disparity are produced by a model-guided filtering strategy from multi-frame LiDAR points. This approach broadens the diversity of training data and allows for controlled experimentation with edge cases and rare events, crucial for enhancing the reliability and adaptability of autonomous systems in practical applications.*

Keywords- Autonomous vehicles, Synthetic data, LidarExtreme conditions, AI, GANs, Machine learning, Safety, Training data

I. INTRODUCTION

The proliferation of AVs represents a significant leap in transportation technology. However, the deployment of AVs in real-world scenarios necessitates rigorous training under various environmental conditions. Traditional data collection methods are often insufficient for capturing extreme conditions such as severe weather, nighttime driving, and rare but critical scenarios like accidents or unusual road configurations. This paper investigates the application of AI-generated synthetic data to bridge this gap, ensuring that AVs are well-prepared for all possible driving conditions.

1.1 Background

Autonomous vehicles (AVs) rely heavily on vast amounts of high-quality data for training their perception and decision-making systems. This data is typically gathered through real-world driving experiences, encompassing various scenarios and environments. However, collecting such data, especially for extreme and rare conditions like heavy snow, torrential rain, dense fog, and nighttime driving, presents significant challenges. These conditions are not only difficult to encounterbut also dangerous for manual collection, which makes comprehensive real-world datasets limited and insufficient.

1.2 Problem Statement

The primary challenge in training AVs for extreme conditions lies in the scarcity of real-world data for these scenarios. Extreme conditions are unpredictable and infrequent, making it hard to collect enough diverse examples for robust model training. Additionally, the variability in these conditions requires data that covers a wide range of situations to ensure the AV's reliability and safety in the real world. This scarcity leads to AVs that perform well under normal conditions but fail to generalize in adverse environments, posing significant risks.

1.3 Objectives

The objective of this paper is to explore the use of artificial intelligence (AI) to generate synthetic data tailored for training AVs in extreme conditions. By leveraging AI, we aim to create high-fidelity synthetic datasets that can simulate a variety of adverse scenarios, thus providing the necessary data diversity and volume required for effective AV training.

II. BACKGROUND AND RELATED WORK

2.1 Real-world Datasets

Existing datasets like KITTI, Cityscapes, and DrivingStereo have been pivotal in advancing AV technologies. However, these datasets primarily focus on normal driving conditions with limited representation of extreme scenarios. The KITTI dataset, for instance, offers valuable data for stereo matching and object detection but lacks sufficient coverage of harsh weather or nighttime conditions. The DrivingStereo dataset improves on this by including more diverse scenarios, yet it still falls short in providing comprehensive extreme condition data

2.2 Challenges in Autonomous Vehicle Training

- **Insufficient data for extreme conditions:**

Real-world data collection often lacks diversity in extreme scenarios, making it challenging to train AVs effectively for all potential situations.

- **Safety risks and high costs:**

Collecting data in severe weather or other hazardous conditions can be dangerous and expensive.

- **Limitations of current simulation tools:**

Existing simulation tools may not adequately mimic real-world complexities, leading to gaps in training.

2.2 Synthetic Data Generation Techniques

The use of synthetic data in AV training is not new. Tools such as CARLA and AirSim have been developed to simulate realistic driving environments, enabling researchers to generate synthetic datasets for training and testing AV algorithms. These simulators offer controlled environments where various scenarios, including extreme conditions, can be created. However, the realism of the generated data and its effectiveness in real-world applications remain ongoing challenges. Techniques are;

- **Generative Adversarial Networks (GANs):**

Introduced by Goodfellow et al., GANs consist of a generator that creates synthetic data and a discriminator that evaluates its realism, iteratively improving the quality of generated data.

- **Recent advancements:**

Techniques like Conditional GANs (cGANs) and CycleGANs have shown promise in generating specific scenarios and translating between different domains (e.g., day to night).

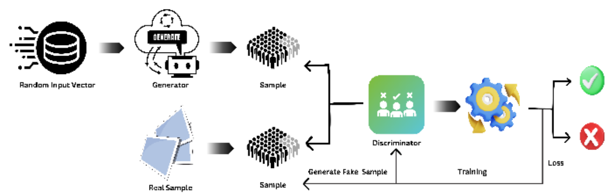
- **Previous applications:**

Synthetic data has been used in other fields such as medical imaging and facial recognition, demonstrating its potential to enhance training datasets.

III. METHOLOGY

Synthetic Data Generation

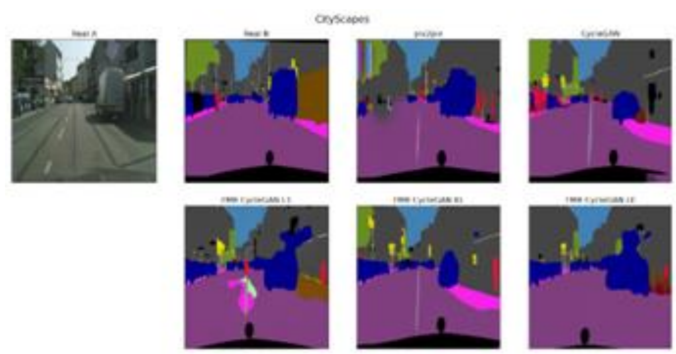
3.1 Generative Adversarial Networks (GANs)



(Fig 1)

GANs consist of two neural networks: the generator and the discriminator. The generator creates synthetic images, while the discriminator evaluates them against real images. Over time, the generator improves to produce highly realistic images.

3.2 Conditional GANs (cGANs) and CycleGANs



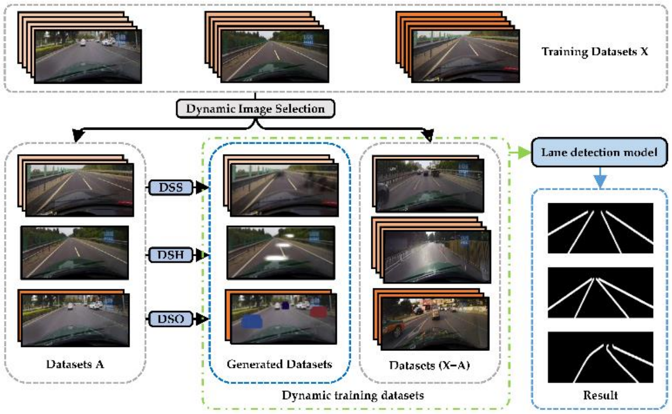
(Fig 2)

- **cGANs:** These networks generate images conditioned on certain inputs, such as generating images of rain or snow given a clear weather image.
- **CycleGANs:** Useful for domain adaptation, CycleGANs can translate images from one domain to another, like converting daytime images to nighttime scenes.

3.3 Data Augmentation and Integration

Augmentation techniques, such as flipping, cropping, and altering brightness, help increase the diversity of the dataset, ensuring models are robust to variations.

Merging Synthetic Data with Real-World Datasets



(Fig 3)

Synthetic data is integrated with real-world datasets to create a comprehensive training set. This involves balancing the dataset to ensure a diverse range of scenarios and validating the augmented dataset through various metrics.

3.31 Balancing the Dataset

1. **Diversity of Scenarios:**

   - **Objective:** Ensure the dataset covers a wide range of driving conditions, including rare and extreme scenarios.
   - **Approach:** Use synthetic data to fill gaps where real-world data is scarce, such as in snowstorms or nighttime conditions.

2. **Proportional Representation:**

   - **Objective:** Maintain a balanced ratio between synthetic and real data to prevent bias.
   - **Approach:** Adjust the volume of synthetic data to complement real-world data without overwhelming it.

3.32 Validation of the Augmented Dataset

1. **Evaluation Metrics:**

   - **Accuracy:** Measure how well models predict or classify objects in diverse conditions.
   - **Robustness:** Test the model's ability to generalize across unseen scenarios.
   - **Consistency:** Check for stable performance across different environments and conditions.

2. **Realism Assessment:**

   - **Visual Inspection:** Conduct manual reviews to ensure synthetic images are indistinguishable from real images.
   - **Perceptual Studies:** Use human evaluators to assess the realism of synthetic data.

3. **Cross-Validation:**

   - **Approach:** Split the dataset into training and testing subsets, ensuring each contains a mix of real and synthetic data.
   - **Objective:** Validate that models trained on this mixed dataset perform well on both real-world and synthetic scenarios.

3.4 Validation and Quality Assurance

Ensuring the quality and effectiveness of synthetic data is crucial:

- **Validation Metrics**: New metrics are proposed to evaluate the synthetic data, focusing on realism, diversity, and relevance to extreme conditions. Metrics such as Mean Absolute Error (MAE) and Intersection over Union (IoU) are used to assess the accuracy and completeness of the generated data.
- **Human-in-the-Loop**: Incorporating human feedback into the validation process helps refine the synthetic data. Experts review the generated scenarios, providing insights and suggestions for improvement, which are then used to further enhance the AI models.

IV. DATA CONSTRUCTION

4.1 Data Acquisition

The data acquisition system comprises multiple color cameras (Basler ACA1920-40GC), a 3D laser scanner (Velodyne HDL-64E S3), and a GPS/IMU navigation system (OXTS RT3003G), all mounted on an SUV. Two cameras are arranged as stereo pairs with a 54 cm baseline and a 50° field of view, while the LiDAR is positioned behind the central camera, and the GPS/IMU system is located at the rear. Accurate spatial alignment between the cameras and LiDAR is achieved through calibration, and timesynchronization is maintained using GPS clocks.

The dataset was constructed using an SUV equipped with multiple cameras, a 3D laser scanner, and a GPS/IMU navigation system. The cameras were used to capture stereo images, and the LiDAR provided point clouds for disparity label generation. The stereo pair had a baseline distance of 54 cm and a field of view of 50 degrees. The dataset includes over 1 million frames, from which 174,437 frames are used for training and 7,751 frames for testing. Disparity labels were generated by integrating point clouds from multiple frames to increase the number of valid pixels. The quality of the labels was further enhanced using a model-guided filtering strategy.

The LiDAR point clouds acquired are sparse, especially at greater distances. When points from a single frame are projected onto an image plane, only a few pixels in the image contain valid values. To increase the number of valid pixels in a single frame, we integrate and fuse point clouds from adjacent frames, as shown in Fig. 3. The data fusion process can be expressed as...

$$C_k^f = \sum_{i=-m}^{n} C_{(k+i)}^s \times T_{(k+i)\to k}, \tag{1}$$

According to calibration matrices, each LiDAR point

$$p^c = P_{rect} \times R_{rect} \times (R|T)_{l\to c} \times p^l, \tag{2}$$

where the P_{rect} is the 3 × 4 intrinsic matrix of left referenced camera, and the R_{rect} is the 4 × 4 calibrated matrix between stereo cameras.

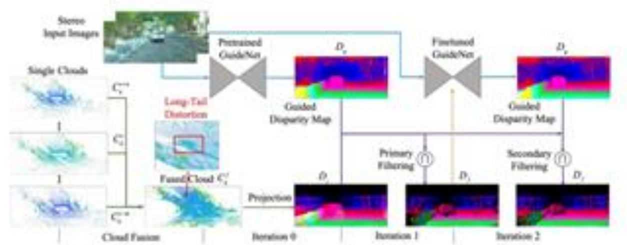
4.2Model-Guided Filtering

Model-guided filtering with LiDAR is an advanced method for producing high-quality disparity labels, crucial for training autonomous systems. This process involves using stereo input images and LiDAR data to create precise and diverse training datasets. The approach starts with cloud fusion, where multiple LiDAR frames are combined to form a comprehensive point cloud. This fused data undergoes a projection step to align it with stereo images, creating an initial disparity map.

GuideNet, a specialized model, refines this map through iterative filtering. Initially, a pretrained version of GuideNet generates a guided disparity map, which undergoes primary filtering to remove inaccuracies. The finetuned GuideNet then applies secondary filtering, further enhancing label quality. This iterative process addresses issues like long-tail distortion, improving the representation of complex environments. By leveraging LiDAR's depth perception, the approach ensures that training data is both accurate and diverse, enhancing model performance across varied conditions. This strategy ultimately leads to more robust and adaptable autonomous systems capable of navigating challenging scenarios.

$$D_f = D_c \cap D_g,$$
$$d_c \cap d_g = \begin{cases} 1 & |d_c - d_g| \le \delta_d \\ 0 & |d_c - d_g| > \delta_d \end{cases}, \tag{3}$$

One of the significant challenges in LiDAR data processing is long-tail distortion, where dynamic objects introduce errors in disparity maps. The model-guided filtering process effectively addresses this issue by iteratively refining the disparity labels. By ensuring consistent and accurate data, this method enhances the overall reliability of training datasets.



(Fig 4)

LiDAR's depth perception capabilities are instrumental in creating detailed environmental representations. This ability allows the system to capture complex scenes accurately, even in challenging conditions such as varying light and weather. The integration of precise depth information improves the training process, making autonomous systems more robust and adaptable to real-world scenarios. By employing model-guided filtering with LiDAR, the training datasets for autonomous systems become significantly more accurate and diverse. This approach not only improves the system's ability to navigate complex environments but also enhances its adaptability to unforeseen challenges. Ultimately, this leads to safer and more reliable autonomous vehicles capable of operating effectively in diverse conditions.

V. EVALUATION METRICES

To accurately assess the performance of stereo matching methods, two primary evaluation metrics—distance-aware metrics and semantic-aware metrics—are employed. These metrics offer a comprehensive evaluation of how well the system performs across different conditions and object types, ensuring that the model's predictions are reliable in real-world scenarios.

5.1 Distance-Aware Metrics

Distance-aware metrics are specifically designed to evaluate stereo matching performance across varying distances, addressing the challenge of maintaining accuracy in diverse range conditions. These metrics are crucial because the accuracy of disparity predictions can significantly vary depending on the proximity of objects to the sensor, especially in real-world driving environments where distances can range from a few meters to hundreds of meters.

Key Metrics:

1. **Absolute Relative Difference (ARD):**
    - **Definition:** ARD measures the relative error between the predicted disparity values and the ground-truth values within specific depth intervals. It is calculated as the absolute difference between predicted and actual disparities, normalized by the actual disparity.
    - **Purpose:** ARD provides a detailed evaluation of stereo matching performance at different distances, ensuring that the model's accuracy is assessed not just at a single average value but across short, middle, and long-range distances. This granular approach highlights how the model performs when objects are near versus when they are far, which is essential for applications like autonomous driving, where precise distance measurement is critical for safety.
2. **Global Difference (GD):**
    - **Definition:** GD is an aggregate metric that sums up the ARD values across all defined distance intervals, providing a holistic measure of the model's performance across varying ranges.
    - **Purpose:** By aggregating errors from different distance intervals, GD offers a comprehensive overview of the stereo matching system's overall accuracy. This metric helps identify any biases the model might have toward specific ranges, such as performing better at short ranges but poorly at longer distances, thus guiding further model adjustments and training.

5.2 Semantic-Aware Metrics

Semantic-aware metrics focus on the accuracy of stereo matching in recognizing and representing various object categories in complex scenes, such as those encountered in autonomous driving. These metrics are essential because different object types—like vehicles, pedestrians, or road signs—pose unique challenges for disparity estimation. Evaluating performance across these categories ensures the model is robust and reliable across a wide array of real-world objects.

Key Metric:

1. **Matching Rate (MR):**
    - **Definition:** MR quantifies the percentage of correctly matched disparity values for objects belonging to specific semantic categories. It evaluates the accuracy of disparity predictions for each object type, such as cars, pedestrians, cyclists, or static obstacles.
    - **Purpose:** By calculating MR for various object categories, semantic-aware metrics provide insights into how well the stereo matching system performs in accurately representing each type of object in the scene. For instance, accurate disparity estimation of vehicles is crucial for maintaining safe distances, while precise pedestrian recognition ensures effective obstacle avoidance. This metric helps developers fine-tune models to improve performance in critical driving scenarios.

Together, distance-aware and semantic-aware metrics offer a nuanced evaluation framework that goes beyond simple accuracy measurements. They help identify specific strengths and weaknesses in the stereo matching process, allowing for targeted improvements. For autonomous driving systems, these metrics are vital for ensuring that models can accurately interpret the complex and dynamic environments they navigate, ultimately leading to safer and more effective operations.

V. EXPERIMENTS

The experimental section provides a comprehensive analysis of the performance of the proposed methods, including the GuideNet model, against established benchmarks. This analysis is conducted through a detailed comparison of datasets and evaluation metrics, highlighting the strengths of the new approaches introduced in this work. By assessing the efficiency and accuracy of disparity predictions, the experiments underscore the effectiveness of the DrivingStereo dataset and the novel evaluation metrics in diverse driving scenarios.

5.1 Dataset and Method Comparison

5.1.1. Overview of the DrivingStereo Dataset:

The DrivingStereo dataset is a large-scale, high-quality stereo dataset specifically designed for autonomous driving applications. It provides a diverse range of driving scenarios with challenging conditions, such as varying lighting, weather, and occlusions, making it a robust benchmark for testing stereo matching algorithms. Compared to traditional datasets, DrivingStereo offers richer semantic information and higher-resolution disparity maps, which are crucial for accurate depth estimation in real-world environments.

5.1.2. Comparison with Existing Datasets:

- **KITTI Dataset:** While KITTI has been widely used for stereo matching tasks, it is limited in terms of scene diversity and semantic complexity. The DrivingStereo dataset expands on these aspects by providing more varied driving scenarios, including urban, suburban, and highway settings. Additionally, it includes a broader range of weather conditions and times of day, such as fog, rain, and night-time scenes, which are underrepresented in KITTI.

- **Middlebury Dataset:** Middlebury focuses on indoor and less dynamic scenes with high-quality ground truth but lacks the scale and contextual variety required for autonomous driving tasks. In contrast, DrivingStereo captures the complexities of outdoor driving environments, providing a more relevant testbed for evaluating stereo matching methods in the context of real-world automotive applications.

5.1.3. Method Comparison and Performance Evaluation:

The experiments also compare the performance of various stereo matching models using the DrivingStereo dataset and novel evaluation metrics, emphasizing the strengths of the GuideNet model.

- **GuideNet Model:**GuideNet, used for generating guided disparity maps, plays a pivotal role in improving stereo matching accuracy. It utilizes pre-trained and fine-tuned versions to progressively refine disparity predictions, which is a critical advantage in accurately depicting depth in complex driving scenarios.

- **Model Efficiency and Accuracy:** Compared to existing methods, such as traditional CNN-based stereo matching models, GuideNet shows superior efficiency and accuracy in generating disparity maps. This is largely due to its architecture, which integrates a multi-stage filtering approach that iteratively refines predictions, significantly reducing errors in disparity estimation across both short and long ranges.

5.1.4. Evaluation Metrics Impact:

The adoption of distance-aware and semantic-aware metrics provides a more detailed and relevant assessment of stereo matching performance compared to traditional evaluation methods. These metrics allow for an in-depth analysis of how well stereo matching models perform across varying distances and object categories, highlighting areas where traditional models may falter.

- **Distance-Aware Metrics Performance:** The experiments reveal that the distance-aware metrics used in conjunction with the DrivingStereo dataset help identify specific performance gains in short, middle, and long-distance disparity estimations. This level of detail is crucial for refining models to handle diverse driving conditions more effectively.

- **Semantic-Aware Metrics Performance:** By evaluating stereo matching accuracy across different semantic categories (e.g., vehicles, pedestrians), the experiments demonstrate that GuideNet and the DrivingStereo dataset better capture critical features needed for safe and reliable autonomous driving. This leads to more accurate object recognition and distance estimation, enhancing the overall system performance in real-world scenarios.

5.1.5. Key Findings and Advantages:

The experimental results highlight several advantages of the DrivingStereo dataset and the GuideNet model:

- **Improved Disparity Accuracy:**GuideNet's guided approach significantly reduces long-tail distortions in disparity predictions, resulting in more accurate and reliable depth maps across a wide range of conditions.

- **Enhanced Performance in Diverse Scenarios:** The DrivingStereo dataset's diverse scenes allow for more rigorous testing and validation, showing clear improvements in handling complex driving environments compared to traditional datasets

.

- **Efficient Model Training and Evaluation:** The combination of high-quality data and targeted evaluation metrics supports efficient model training, leading to faster convergence and better overall model performance.

VI. CONCLUSION

The experiments conclusively demonstrate that the DrivingStereo dataset, along with GuideNet and the new evaluation metrics, sets a new standard in stereo matching performance for autonomous driving. By addressing the limitations of existing datasets and methods, this approach provides a more comprehensive and accurate assessment framework, leading to significant advancements in stereo data quality, model robustness, and real-world application effectiveness.

REFERENCES

[1] G. Yang, X. Song, C. Huang, Z. Deng, J. Shi and B. Zhou, "DrivingStereo: A Large-Scale Dataset for Stereo Matching in Autonomous Driving Scenarios," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 899-908, doi: 10.1109/CVPR.2019.00099.

[2] L. Li, K. N. Ismail, H. P. H. Shum and T. P. Breckon, "DurLAR: A High-Fidelity 128-Channel LiDAR Dataset with Panoramic Ambient and Reflectivity Imagery for Multi-Modal Autonomous Driving Applications," *2021 International Conference on 3D Vision (3DV)*, London, United Kingdom, 2021, pp. 1227-1237, doi: 10.1109/3DV53792.2021.00130.

[3] D. Kondermann *et al*., "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving," *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, 2016, pp. 19-28, doi: 10.1109/CVPRW.2016.10.

[4] D. Kondermann *et al*., "The HCI Benchmark Suite: Stereo and Flow Ground Truth with Uncertainties for Urban Autonomous Driving," *2016 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Las Vegas, NV, USA, 2016, pp. 19-28, doi: 10.1109/CVPRW.2016.10.

[5] B. Liang *et al*., "Real-Time Stereo Image Depth Estimation Network With Group-Wise L1 Distance for Edge Devices Towards Autonomous Driving," in *IEEE Transactions on Vehicular Technology*, vol. 72, no. 11, pp. 13917-13928, Nov. 2023, doi: 10.1109/TVT.2023.3284011

[6] Revisiting Stereo Depth Estimation From a Sequence-to-Sequence Perspective with Transformers Zhaoshuo Li, Xingtong Liu, Nathan Drenkow, Andy Ding, Francis X. Creighton, Russell H. Taylor, Mathias Unberathhttps://doi.org/10.48550/arXiv.2011.02910

[7] Stereo R-CNN based 3D Object Detection for Autonomous Driving Peiliang Li, Xiaozhi Chen, Shaojie Shen https://doi.org/10.48550/arXiv.1902.09738

[8] Xiao Song, Xu Zhao, Hanwen Hu, and Liangji Fang. Edgestereo: A context integrated residual pyramid network for stereo matching. In ACCV, 2018

[9] P. Chen et al., "ECMD: An Event-Centric Multisensory Driving Dataset for SLAM," in IEEE Transactions on Intelligent Vehicles, vol. 9, no. 1, pp. 407-416, Jan. 2024, doi: 10.1109/TIV.2023.3339144.

[10] Y. Zhang, L. Wang, K. Li, Z. Fu and Y. Guo, "SLFNet: A Stereo and LiDAR Fusion Network for Depth Completion," in *IEEE Robotics and Automation Letters*, vol. 7, no. 4, pp. 10605-10612, Oct. 2022, doi: 10.1109/LRA.2022.3191235