CleanMAP: Distilling Multimodal LLMs for Confidence-Driven Crowdsourced HD Map Updates

Supplementary Material

7. Factors Affecting Adaptive Confidence Score Calculation

While calculating the Adaptive Confidence Score, several factors must be considered, including key parameters affecting lane line visibility and the importance of these factors. The final confidence score of an input image is determined by dynamically assigning weights based on environmental conditions and their impact on lane visibility.

7.1. Designing Specific Parameters to Calculate Confidence Score

To ensure the reliability and accuracy of HD map updates, we define a set of key parameters that impact image quality and influence the visibility of lane markings. These parameters form the basis for the MLLM-based scoring system, where it assigns an individual score between 0 and 10 to each parameter based on its impact on lane visibility. The key parameters includes:

7.1.1. Blur

Image clarity is essential for detecting lane markings and road features. Different types of blur that affect visibility include:

- Daytime Blur: Caused by motion or camera focus issues.
- Nighttime Blur: Often due to low light or motion in poorly illuminated areas.
- Streetlight-Induced Blur: Occurs when lane markings are obscured by artificial light sources at night.

7.1.2. Illumination

Strong lighting variations can impact image clarity, including:

- Strong Sunshine or Shadows: Excessive brightness or deep shadows obscuring road features.
- Reflections or Glare: From reflective road surfaces or vehicles.
- Darkness: Low-light conditions where lane markings become less visible.

7.1.3. Weather Conditions

Environmental conditions can obscure road markings and reduce visibility:

- Rain, Snow, and Fog: Adverse weather conditions that diminish lane visibility.
- Sandstorms: In desert regions, sandstorms can reduce visibility to near zero, affecting map updates.

7.1.4. Lane Line Degradation

Over time, lane markings may wear out and become unclear or invisible, making them unreliable for HD map updates.

7.1.5. Obstacles Covering Lane Lines

Vehicles, debris, or objects on the road may obstruct lane markings, making it difficult to assess road conditions accurately.

7.1.6. Lane Line Visibility

The overall visibility of lane markings in an image directly impacts its usability for HD map updates.

7.2. Designing Specific Task-Guided Questions/Prompts

To ensure the model accurately assesses image quality, a Task-Guided Instruction Prompting system is implemented. This system guides the model through structured prompts that focus on critical aspects such as lane line visibility, weather conditions, obstacles, and different types of blur. By directing the model's attention to relevant factors, the resulting confidence scores remain contextually accurate.

7.2.1. Task-Guided Questions for Image Evaluation

Each prompt is designed to assess a specific aspect of the image, ensuring structured evaluation:

- Question 1: Detailed Scene Description Prompt "Provide a detailed description of the scene in the image, focusing on lane line visibility, the impact of vehicles or obstacles, weather conditions, and any other factors affecting clarity."
- Question 2: Daytime Blur Prompt "Rating blurred image during daytime: [Score 0-10] - Rate the overall image clarity/sharpness on a scale of 0-10, where 10 is extremely blurry and 0 is tack sharp."
- Question 3: Nighttime Blur Prompt "Rating blurred image during nighttime: [Score 0-10] - Rate the image clarity on a scale of 0-10, considering lane line visibility."
- Question 4: Streetlight Blur at Night Prompt "Rating blurred lane lines due to Street Lights at Night: [Score 0-10] - Rate the clarity of lane lines, where 10 is extremely blurred and 0 is perfectly sharp."
- Question 5: Lane Line Invisibility due to Illumination Prompt "Rating Lane Lines Invisibility due to Illumination (strong sunshine/shadows/darkness): [Score 0-10] - Rate how invisible lane lines are due to strong illumination effects."

- Question 6: Invisibility due to Fog Prompt ”Rating Lane Lines Invisibility due to Fog: [Score 0-10] - Rate the extent to which lane lines are obscured by fog.”
- Question 7: Invisibility due to Rain Prompt ”Rating Lane Lines Invisibility due to Rain: [Score 0-10] - Rate how blurred lane lines are due to rain.”
- Question 8: Invisibility due to Snow Prompt ”Rating Lane Lines Invisibility due to Snow: [Score 0-10] - Rate how snow obscures lane lines.”
- Question 9: Invisibility due to Sandstorm Prompt ”Rating Lane Lines Invisibility due to Sandstorm: [Score 0-10] - Rate how blurred lane lines are due to sand.”
- Question 10: Lane Line Degradation Prompt ”Rate the condition of lane lines on a scale of 0 to 10, where 0 is completely worn off and 10 is perfectly clear.”
- Question 11: Vehicles Obstructing Lane Lines Prompt ”Rate the visibility of lane lines blocked by vehicles, where 10 is fully blocked and 0 is fully visible.”
- Question 12: Overall Lane and Lane Marking Visibility Prompt ”Rate the overall visibility of the lanes and lane markings in the image on a scale of 0-10, where 10 means they are clearly visible, and 0 means they are completely invisible.”

7.3. Efficient Selection of Parameters Based on Context

Evaluating all parameters in every scenario may lead to hallucination, where the model assigns arbitrary or inaccurate scores. To mitigate this, parameter selection is dynamically adjusted based on environmental conditions:

- In clear, sunny weather, irrelevant parameters such as rain, snow, and fog are omitted to prevent unnecessary noise in the scoring process.
- In adverse weather conditions such as heavy rain, snow, or fog, the weights of these factors are increased due to their significant impact on image clarity. Conversely, parameters such as illumination and streetlight blur, which become less relevant, are weighted lower.

This adaptive parameter selection optimizes the model's focus on relevant factors, reducing the risk of hallucinations and ensuring confidence scores remain accurate.

7.4. Importance of Weight Assignment

Each parameter affects image quality differently depending on environmental conditions. To ensure accurate confidence score calculations, parameter weights are dynamically adjusted:

- In images collected during rain, snow, or fog, higher weights are assigned to weather-related parameters as these conditions obscure lane markings.
- In clear conditions, illumination-related factors such as reflections, shadows, and glare from the sun or streetlights are weighted more heavily.
- Degradation and obstacle-related parameters are assigned significant weights in all conditions, as they consistently affect lane marking detection.

The model dynamically adjusts these weights to maintain accuracy by ignoring irrelevant parameters in specific conditions. For example, fog-related parameters in clear weather are either assigned low weights or ignored to prevent unnecessary confusion in the model's reasoning process.

Therefore, by integrating adaptive parameter weighting, task-guided instruction prompting, and context-aware parameter selection, the confidence scoring model ensures precise and reliable assessments. This approach enhances HD map update accuracy by filtering unreliable data while preserving essential information.

8. Diverse Data Collection and Annotation for MLLM-Driven Confidence Scoring

8.1. Diverse Data Collection for Training

To ensure that the model can accurately assess data quality across a wide range of conditions, a small but diverse dataset was collected, consisting of both real-world and synthetic images, as shown main paper Figure 3. This dataset includes images captured from connected and automated vehicles (CAVs) under various environmental conditions, as well as handcrafted images designed to simulate specific adverse scenarios.

The dataset is composed of two primary sources:

- **Online Crowdsourced Data:** Real-world images were gathered from vehicles operating in urban and rural environments under varying weather conditions. These images include those captured in daylight and nighttime, as well as during adverse weather such as rain, snow, fog, and dust storms. This diverse set ensures that the model learns to handle a broad range of conditions that may degrade data quality.
- **Synthetic Images:** To supplement real-world data, synthetic images were generated to simulate extreme conditions such as heavy rain, dense fog, and severe illumination effects, including reflections, shadows, and glare from headlights and streetlights. These synthetic images enable the model to generalize to rare but critical conditions that are essential for robust HD map updates.

The dataset comprises approximately 1,000 images, including blurred images from both daytime and nighttime scenarios. This curated dataset enables the model to learn how different factors, such as blur, lighting, and weather, affect data quality. By exposing the model to a wide range of adverse conditions, the training process ensures accurate real-time data quality assessment under real-world constraints.

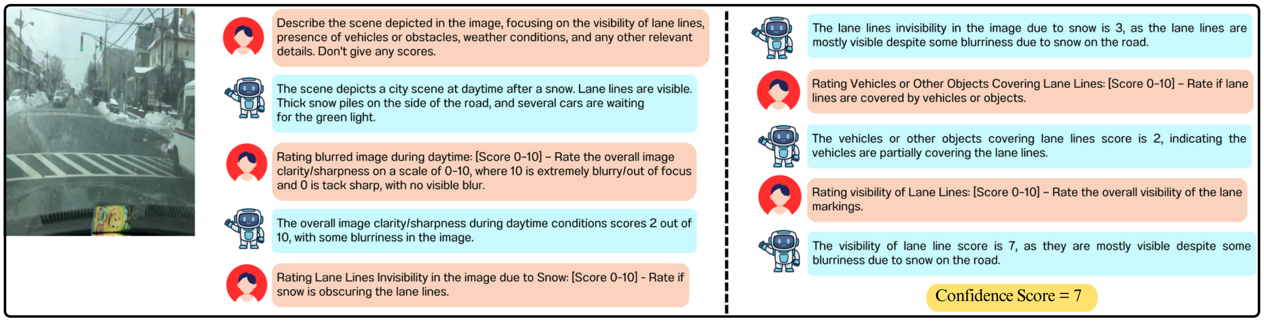


Figure 7. Scenario 1: Snow Conditions with Minor Blur.

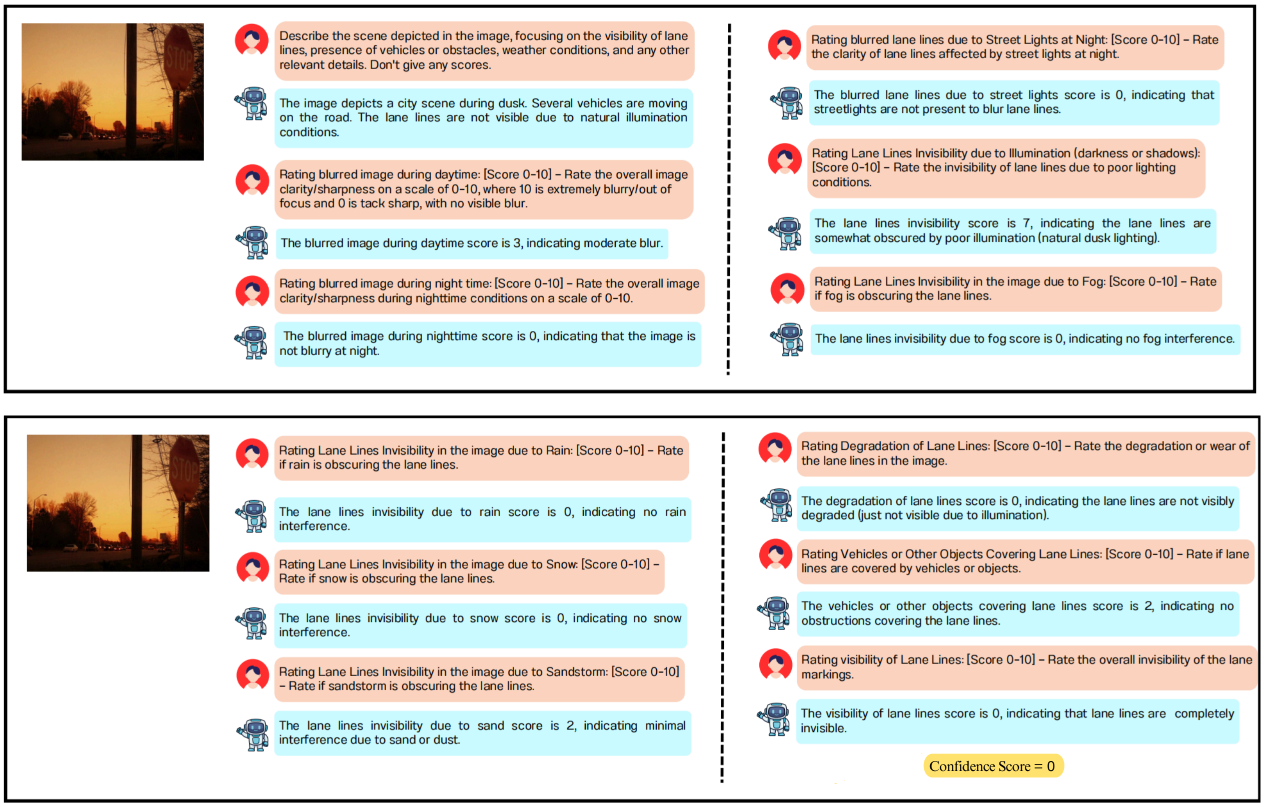


Figure 8. Scenario 2: Dusk Scene with Poor Illumination. The model evaluates lane line visibility under poor lighting conditions, correctly filtering out irrelevant factors while highlighting the impact of natural illumination.

8.2. Accurate Annotation of Collected Data

The collected data is meticulously annotated based on predefined parameters that measure image quality. Each image is manually scored across multiple factors affecting lane line visibility and overall road conditions. These annotations provide a strong baseline for the model during training.

Manual annotation enables the model to learn the relationship between visual cues and external conditions such as blur, rain, snow, and illumination, which impact road visibility to varying degrees. For instance, lane markings may become nearly invisible in dense fog but remain relatively clear in mild rain. Accurate annotations allow the model to capture these nuances by assigning well-defined scores for each factor. Without structured annotations, the model would struggle to interpret how different conditions influence lane visibility and image clarity.

For each image, multiple parameters are assessed and scored to capture how different conditions affect image quality and lane line visibility:

- **Weather Conditions (Fog, Rain, Snow, Sandstorm):**

Natural elements significantly impact lane line visibility. Each weather condition is scored on a scale (e.g., 0-10) to reflect its severity in the image. Precise annotation ensures the model can appropriately adjust confidence scores for images affected by these conditions.

- **Blur (Daytime and Nighttime):** Blur can result from camera motion, poor focus, or adverse lighting. Since its causes and impacts vary between daytime and nighttime, separate annotations for each condition are necessary to ensure proper learning.
- **Illumination (Sunshine, Shadows, Darkness):** Strong sunshine, deep shadows, or nighttime darkness can obscure lane markings, making them difficult to detect. Accurate annotation of illumination levels ensures the model correctly evaluates when lighting conditions affect lane visibility, adjusting the image's usability score accordingly.
- **Degradation of Lane Lines:** Over time, lane markings degrade, becoming unclear or invisible. Annotating the condition of lane markings is essential, as it directly impacts the model's ability to evaluate road geometry and lane detection reliability.
- **Obstacles Covering Lane Lines (Vehicles, Debris):** Vehicles, debris, or other objects that obscure lane markings should be annotated. The extent to which lane lines are blocked directly affects the accuracy of HD map updates.
- **Overall Lane Line Visibility:** The overall visibility of lane markings is the most critical factor in determining the usability of an image for HD map updates. This parameter encapsulates how all the aforementioned conditions collectively impact lane detection.

Thus integrating diverse real-world and synthetic data, along with precise manual annotations, the dataset ensures that the model learns to assess data quality across a wide range of challenging conditions. The structured annotation process enables the model to differentiate between various environmental factors, ultimately improving confidence scoring for HD map updates.

9. Additional Qualitative Analysis of MLLM-driven Confidence Scoring

9.1. Scenario-Based Evaluations

9.1.1. Scenario 1: Snow Conditions with Minor Blur

The scenario in Figure 7 shows a post-snow city scene where thick snow piles are visible on the sides of the road. The lane lines are mostly visible, but there is some blur in the daytime conditions. The MLLM was able to capture the following aspects correctly:

- **Blur Detection:** The model rated the image clarity as 2/10, indicating a minor blur that slightly affects lane visibility.
- **Snow Effect:** The model correctly identified snow interference but rated the lane line visibility at 3/10, indicating the lane lines are mostly visible but slightly obscured by snow.
- **Obstruction by Vehicles:** The model recognized that some vehicles partially obstruct the lane lines ahead, giving a score of 2/10 for vehicles covering the lane lines.
- **Overall Lane Line Visibility:** The model rated lane visibility at 7/10, implying that the image provides sufficient clarity for most parts of the lane, despite the snow and slight blur.

The MLLM performed accurately in identifying the key visibility obstructions, particularly the snow and blurriness. It successfully excluded irrelevant factors like fog and nighttime blur, making this a strong example of efficient MLLM application. The confidence score of 7 reflects that the image is still usable for map updates, although the visibility could be affected by snow.

This section provides additional qualitative analysis of of the MLLM-driven confidence-scoring model, including scenario-based evaluations.

We present qualitative assessments across different environmental conditions, highlighting CleanMAP's ability to accurately identify lane visibility challenges and compute confidence scores.

9.1.2. Scenario 2: Dusk Scene with Poor Illumination

In this scenario, illustrated in Figure 8, the image was captured at dusk, where poor illumination significantly affects lane line visibility. The model identified key factors impacting visibility:

- **Illumination Issues:** The model rated lane line invisibility at 6/10, indicating moderate obstruction due to natural lighting conditions.
- **Blur Detection:** The daytime blur score was 2/10, suggesting slight distortion in the image.
- **Environmental Factors:** As no snow, fog, or rain were present, these parameters were correctly assigned a score of 0.
- **Obstruction:** No obstructions from vehicles or objects were detected, receiving a 0 score.

This scenario highlights MLLM's capacity to accurately assess illumination-based visibility challenges while effectively filtering out irrelevant conditions. The confidence score of 1.8 confirms that the image is of low quality for map updates.

9.1.3. Scenario 3: Bright Sunlight Causing Glare

This scenario, depicted in Figure 9, captures a high-illumination road scene where strong sunlight causes glare and partial obstruction by a truck. The model effectively recognized:

- **Illumination Problems:** The model scored lane line invisibility at 7/10, attributing poor visibility to intense glare.

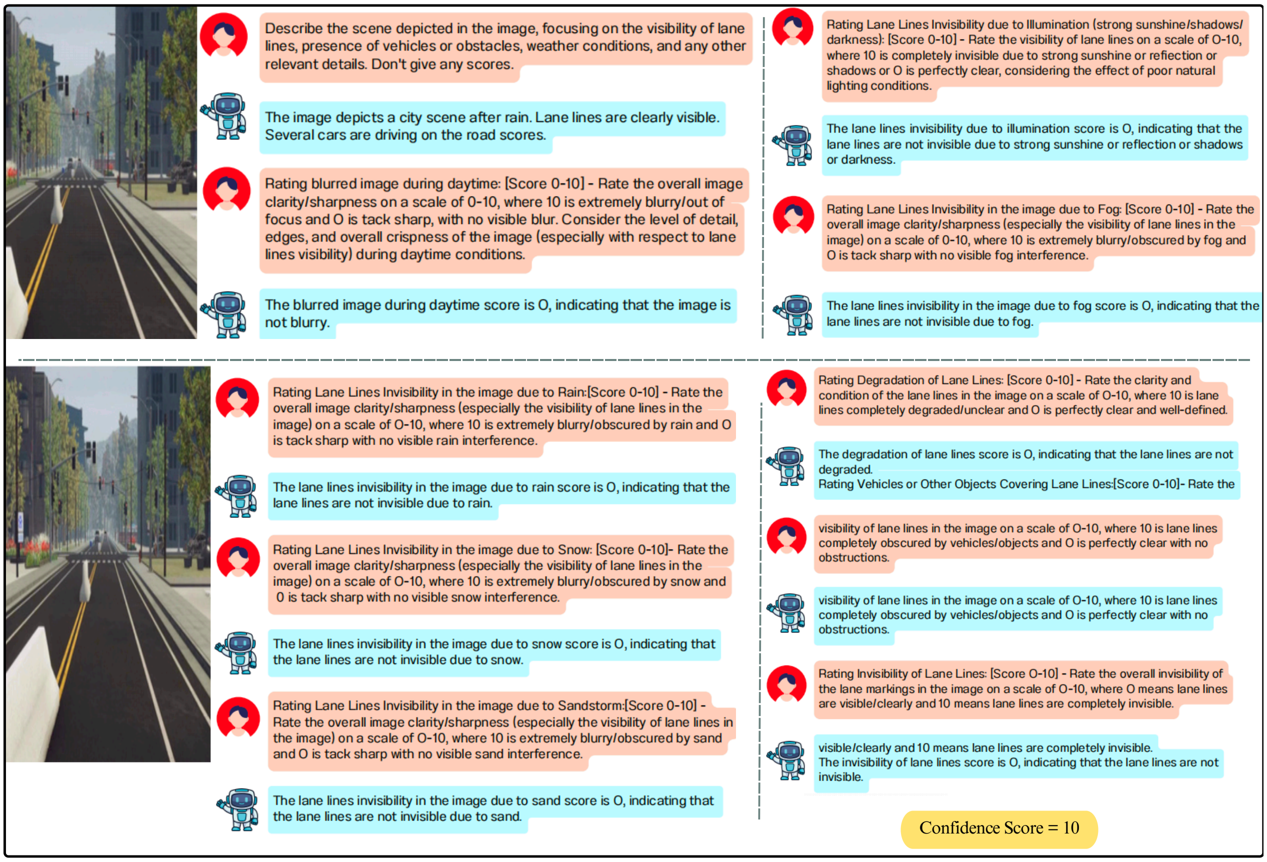


Figure 9. Scenario 3: Bright Sunlight Causing Glare. The model correctly identifies intense glare and obstruction from a truck, demonstrating its efficiency in visibility scoring.

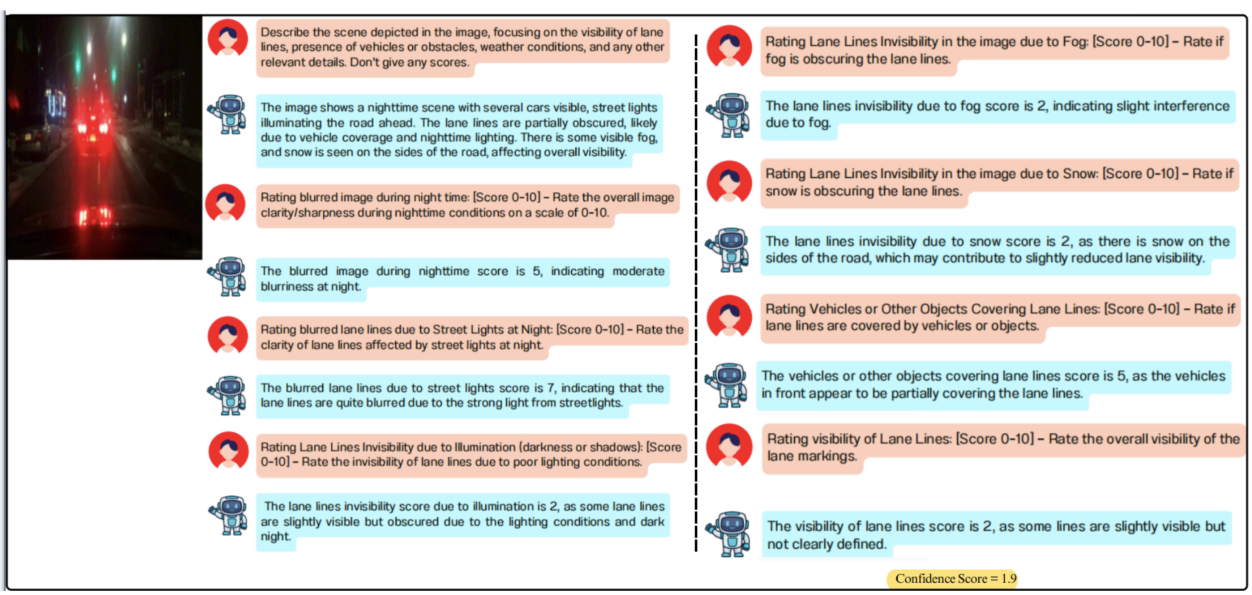


Figure 10. Scenario 4: Snow-Covered Lane with Night Illumination.

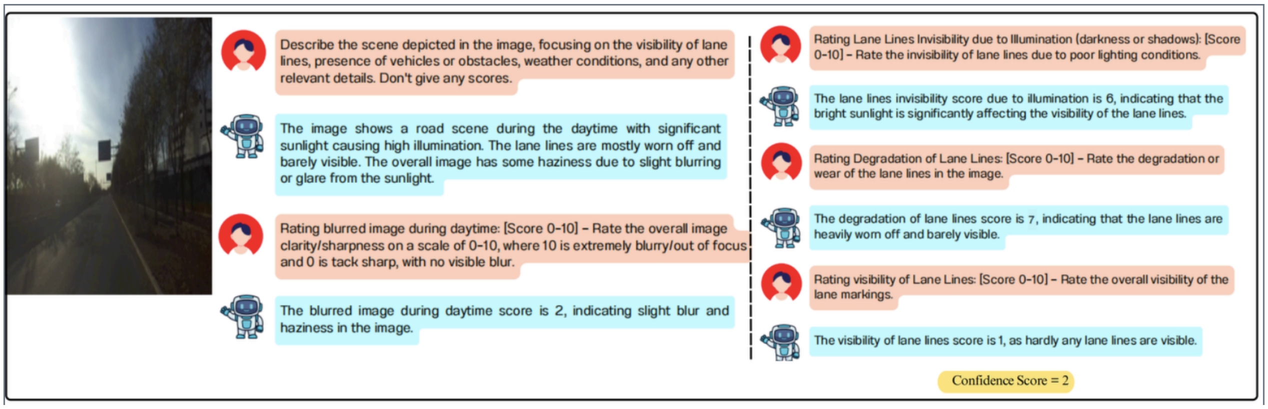


Figure 11. Scenario 5: Daytime Scene with Glare and Degraded Lane Lines.

- **Daytime Blur:** A blur score of 2/10 was assigned, indicating slight haziness due to sunlight.
- **Obstruction by Vehicles:** The truck partially obstructed lane markings, with a vehicle obstruction score of 4/10.
- **Lane Visibility:** The lane lines received a low visibility score of 3/10, confirming the compounded impact of glare and obstruction.

This scenario illustrates MLLM's ability to distinguish between different environmental factors. The confidence score of 0.4 confirms that the image is unsuitable for HD map updates due to major glare issues.

9.1.4. Scenario 4: Snow-Covered Lane with Night Illumination

Figure 10 presents a nighttime scenario where lane visibility is affected by snow accumulation and artificial lighting from streetlights and vehicles. The model's assessment includes:

- **Nighttime Blur:** Moderate blurring due to night conditions was rated at 5/10.
- **Streetlight Effects:** The model assigned a high score of 7/10 to streetlight-induced visibility degradation.
- **Environmental Factors:** Fog and snow were correctly identified with scores of 2/10 each.
- **Obstruction by Vehicles:** Vehicles partially covering lane lines were scored at 5/10.
- **Lane Visibility:** The lane lines received a poor visibility rating of 2/10, confirming significant degradation.

This scenario demonstrates MLLM's ability to identify multiple factors affecting visibility. The confidence score of 1.9 indicates that the image is not ideal for map updates but confirms the model's effectiveness in parameter selection.

9.1.5. Scenario 5: Daytime Scene with Glare and Degraded Lane Lines.

This scenario, as depicted in Figure 11, shows a road during the daytime with strong sunlight affecting the visibility of lane lines:

- **Daytime Blur:** The model assigns a blur score of 2/10, indicating slight blur and haziness, which is accurate given the strong sunlight affecting the scene's sharpness.
- **Illumination Problems:** Lane line invisibility due to illumination was rated at 6/10, suggesting that bright sunlight significantly affected lane line clarity.
- **Degradation of Lane Lines:** The model identified significant wear and tear on the lane lines, assigning a score of 7/10 for lane degradation.
- **Lane Visibility:** Lane lines are rated poorly for visibility, receiving a score of 1/10, as they are barely discernible due to both degradation and bright sunlight.

This scenario highlights the model's ability to correctly assess both glare and lane line degradation. The assessment of blur, lane visibility, and the impact of sunlight is accurate, as is the degradation score. The model appropriately ignores irrelevant factors such as rain and fog, which are not present in the image. The low confidence score of 2 reflects the poor overall quality of the image for mapping purposes.

9.2. Optimal Data Selection for HD Map Updates

To ensure optimal data selection for HD map updates, the confidence-driven fusion strategy prioritizes high-quality local maps. The selection process follows these principles:

- **High-confidence images** are prioritized for inclusion in HD map updates.
- **Dynamic confidence thresholds** are used to avoid excessive filtering and ensure data sufficiency.
- **Environmental adaptability** ensures that the model dynamically adjusts scoring weights based on real-world conditions.

This systematic approach significantly improves the accuracy and reliability of HD maps while maintaining efficient data processing. The supplementary qualitative results confirm CleanMAP's capability to robustly assess and score lane visibility across diverse environmental conditions. The model demonstrates strong adaptability by correctly identifying glare, poor illumination, and snow-related obstructions while filtering out irrelevant conditions. The confidence-driven scoring and data fusion approach ensures that only high-quality images contribute to HD map updates, enhancing reliability in autonomous navigation.

10. Systematic Workflow of MLLM-Driven Confidence-Based HD Map Updates

In HD map updates, integrating local map data from crowdsourced vehicles presents both opportunities and challenges. The objective is to generate a reliable global HD map by fusing individual local maps while ensuring geometric consistency, feature alignment, and positional accuracy.

Let M_{global} represent the global HD map, constructed as the union of multiple local maps M_{local_i} contributed by individual vehicles:

$$M_{\text{global}} = \bigcup_{i=1}^{n} M_{\text{local}_i} \tag{19}$$

where M_{local_i} consists of spatial data points (x_i, y_i, z_i), representing key road features such as lane lines and road boundaries. To fuse these maps, geometric alignment is performed to bring all local maps into a common coordinate system by minimizing deviations in overlapping data points and compensating for sensor inaccuracies and trajectory differences.

Once aligned, the final fusion step involves clustering algorithms to group closely related data points while filtering noise. This structured approach ensures an accurate, up-to-date global HD map that reflects real-time road conditions. By integrating geometric alignment and confidence-driven fusion, the model enhances HD map precision and reliability, making it highly effective for autonomous navigation.

10.1. Optimal Local Maps Selection

The model calculates an average confidence score for each local map, reconstructed from a sequence of timestamped images. Each image is assigned an individual confidence score by the MLLM-based Scoring model, evaluated based on environmental conditions and lane line visibility. These confidence scores are used to rank and organize local maps within specific map links, ensuring that only the most reliable data is utilized for further processing. Local maps with higher confidence scores are given preference for association. This selection process enhances the accuracy and consistency of HD map updates by prioritizing high-confidence local maps while filtering out unreliable data.

10.2. Introducing Changes in Prior Local Maps for Future Lane Line Updates

In real-world scenarios, local map data evolves due to road construction, lane shifts, or infrastructure modifications. To evaluate the effectiveness of the map update process, intentional modifications are introduced in prior local maps, enabling a realistic assessment of how new information is integrated into the existing HD map.

Modifications are performed through three primary tasks:

- **Shifting:** Existing lane lines are shifted in the X and Y directions to simulate lane position changes due to maintenance or expansion. The original lane line is replaced by the shifted one.
- **Deleting:** An entire lane line is removed from the map, representing real-world lane closures or removals.
- **Adding:** New lane lines are introduced between existing ones to simulate road expansion or newly constructed lanes. A new lane is created by calculating the midpoint between two existing lanes with a slight offset to prevent overlap.

Once modifications are applied, the updated local map is saved and compared with the ground truth map to evaluate update accuracy.

Table 8. Definition of HD Map Element Update

Update Task	Prior Map	Fused Local Map	Updated Map
Shifting	Existent	Existent	Existent
Deleting	Existent	Non-existent	Non-existent
Adding	Non-existent	Existent	Existent

Table 8 summarizes how each modification task is performed. Before the update, tasks such as shifting and deleting apply to existing HD map elements. After the update, shifted elements retain their presence with altered positions, deleted elements are removed, and newly added elements are introduced into the HD map from the fused local map.

10.3. Association of Modified and Reconstructed Local Map Data

After selecting the sequences with the highest confidence scores, the model aligns them with the modified local map data using the Iterative Closest Point (ICP) algorithm. This step ensures that the reconstructed local map data points, derived from crowdsourced vehicle-collected image keyframes, are accurately aligned with the modified map data. This alignment facilitates an effective association between the two, ensuring consistency in the HD map update process.

The map association process using the Iterative Closest Point (ICP) algorithm is formulated as an optimization problem to find the optimal transformation that aligns the points in a local map M_{local_i} with those in a subsequent modified map $M_{\text{local}_{i+1}}$. This transformation is represented as the matrix T, which minimizes the alignment error between corresponding points in the two maps:

$$T = \arg\min_{T} \sum_{i=1}^{N} \|M_{\text{local}_i} - T M_{\text{local}_{i+1}}\|^2 \tag{20}$$

where:

- M_{local_i} and $M_{\text{local}_{i+1}}$ represent the sets of points in the local maps before and after modification, respectively.
- T is the transformation matrix, consisting of a rotation matrix R and a translation vector t, which aligns the two maps by minimizing positional error.

The transformation matrix T is expressed as:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{21}$$

where:

- R is the rotation matrix.
- t is the translation vector.

The alignment process determines the optimal R and t by minimizing the discrepancy between corresponding points in the two maps. This ensures that changes in local map data are accurately aligned with the confidence score-based map, maintaining consistency and accuracy in the HD map update. By integrating this association process, the system incorporates the most reliable and up-to-date information into the HD map, ensuring a highly precise representation of the environment.

10.4. Data Fusion for Map Update

The final stage of the HD map update process involves fusing the aligned sequences with the modified local map data points. Clustering algorithms such as DBSCAN (Density-Based Spatial Clustering of Applications with Noise) are employed to fuse the data, ensuring that new and validated information is integrated into the HD map. This step accurately reflects changes in road features, lane lines, and other elements.

Map data collected from crowdsourced vehicles can vary in quality and may include noise or irrelevant information. To address this, DBSCAN is applied to cluster data points from both the associated local maps and the confidence score-based selected map. DBSCAN is particularly effective as it identifies valid clusters (e.g., lane lines and boundaries) while filtering out noise caused by sensor inaccuracies or environmental variations. The fusion of selected maps ensures that the HD map remains up-to-date and accurate.

DBSCAN is defined by two key parameters:

- **Epsilon** (ϵ): The maximum distance between two points for them to be considered part of the same cluster.
- **Min Samples**: The minimum number of points required to form a dense region (cluster).

Let:

- M_{local} represent the local map points, where each point has coordinates (x_i, y_i).
- M_{cs} represent the confidence score-based selected map points with coordinates (x_j, y_j).

DBSCAN is applied to the combined dataset:

$$M_{\text{combined}} = \{M_{\text{local}}, M_{\text{cs}}\} \tag{22}$$

The clustering process is formulated as:

$$C = \text{DBSCAN}(M_{\text{combined}}, \epsilon, \text{min\_samples}) \tag{23}$$

where:

- ϵ is the neighborhood distance parameter that determines the density threshold.
- min_samples is the minimum number of points required to form a cluster.
- C is the set of clusters generated by DBSCAN, with noise points labeled as outliers.

Therefore, by clustering valid map features and filtering out noise, DBSCAN enables the fusion of reliable map data points while discarding outliers. Unique clusters are assigned specific colors, while noise points (if any) are marked in black. This ensures that only the most accurate and up-to-date information is used in HD map updates, maintaining the integrity and precision of the map for autonomous navigation.

11. Additional Information About Real Vehicle Crowdsourced Evaluation Data

11.1. Experiment Setup and Crowdsourced Data Collection

Figure 12. Real time Crowdsourced Data collection Vehicle.

Table 9. Sensors equipped on the Xiaopeng G3 vehicle.

Sensor Type	Model	Parameter
Camera	LI-AR0231-AP0200-GMSL2	1920×1080 @28fps
Lidar	RS-Lidar-32	360°HFOV and 40° VFOV
RTK-GNSS/IMU	NovAtel PP7D-E1	10cm
GNSS	Ublox F9P	10m
Computer	Nuvo-6108GC	Intel i7 + Nvidia 1080

This experiment utilized a real-world vehicle test platform based on the Xiaopeng G3 vehicle, as shown in Figure 12. The vehicle was equipped with multiple sensors, including cameras, commercial-grade GNSS systems, and integrated inertial navigation devices. An onboard industrial-grade computer served as the central processing unit for real-time computations. A detailed list of the sensors and computing platforms used in the Xiaopeng G3 is provided in Table 9.

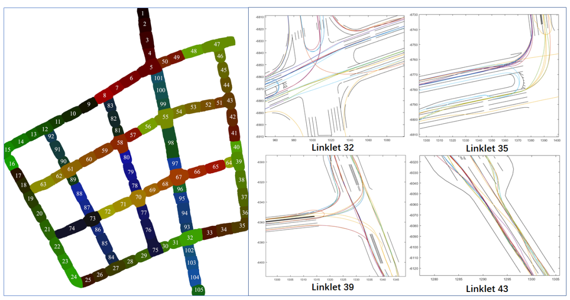
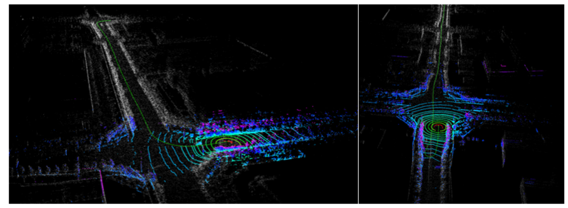


Figure 13. Global Map Representation Composed of Link Areas with Lane Line Coverage.

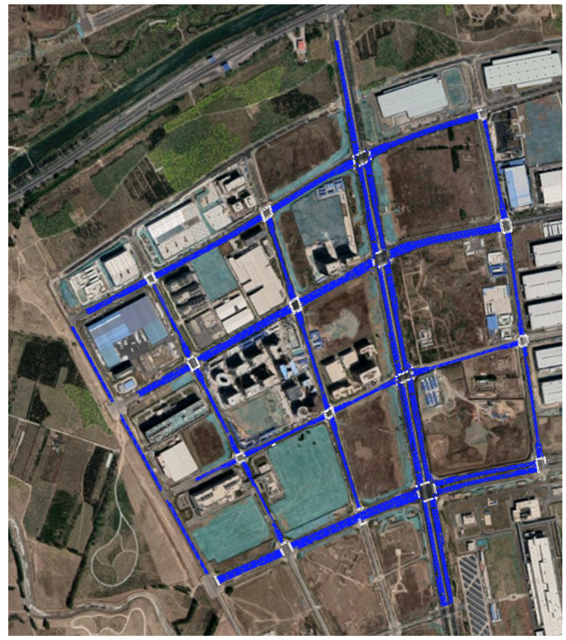
The testing was conducted in an urban environment along roads in the Economic and Technological Development Zone, Daxing District, Beijing. The test area covered approximately 20 kilometers and included diverse urban road structures such as dual lanes, four-lane and six-lane dual carriageways, and multiple intersections with complex lane markings, including solid yellow and white lines, as well as merging and diverging lanes. Figure 13 illustrates the global map of the test area, which was divided into multiple Link Areas, each with specific lane line coverage.

In this area, a high-definition map was generated as the ground truth to evaluate the accuracy of the collected data. The process began with a professional survey vehicle equipped with LiDAR sensors, capturing precise point cloud data of the environment. This data formed the foundation of a high-resolution point cloud map, as illustrated in Figure 14.

Subsequently, essential map features relevant to autonomous driving, such as lane markings, were manually annotated onto the point cloud map. The annotated map provided ground truth data containing road surface elements, ensuring a reliable reference for comparison. Table 10 outlines the specific map features, including the total number of road lines represented as 3D line annotations.



(a) Point cloud map data generation



(b) Annotated result used as ground truth

Figure 14. Generation of Groudtruth Map data.

Table 10. Annotated map features in the experimental data.

Map Elements	Total Number	Annotation Format
Road Lines	1536	3D Line

To simulate real-world conditions, data collection was conducted using a crowdsourced approach, leveraging observations from multiple vehicles at various times and angles. Crowdsourced data provides a diverse range of perspectives, capturing dynamic changes in road conditions. The data collection process spanned eleven months, from October 2022 to September 2023, with data gathered daily between 9 AM and 5 PM to ensure broad temporal diversity. Randomized routes were selected to maximize coverage across the mapped area, allowing the system to capture multiple perspectives of the same location under different conditions.

This approach ensures a comprehensive dataset that accurately reflects variations in road structures, lane visibility, and environmental conditions, contributing to a more reliable and up-to-date HD map.

11.2. Significance of Confidence Score

The confidence score is crucial for determining which data can be reliably integrated into the map update process. Scores range from 0 to 10, with higher scores indicating better data quality. The classification of confidence scores is as follows:

- **Confidence Score ∼2:** These images are of very low quality, often affected by extreme lighting, severe weather conditions such as rain or fog, or obstructions. Such images contribute little to accurate mapping and are generally deemed unreliable.
- **Confidence Score ∼5:** Images in this range are of moderate quality. While they may capture some useful information, they often contain partial obstructions, uneven lighting, or slightly blurred lane markings, limiting their reliability for precise mapping.
- **Confidence Score ∼9:** Images in this range are considered high quality. Captured under ideal environmental conditions, they provide clear visibility of lane markings, traffic signs, and other essential map elements, making them highly reliable for HD map updates.

Table 11. Visual Explanation of what each Score signifies in terms of Image Quality

	Images		
			
Parameters	**MLLM**	**MLLM**	**MLLM**
Blur (Daytime)	1	0	0
Illumination	5	2	1
Degradation	0	0	0
Object	0	1	1
Visibility	3	6	10
Confidence Score	1.8	6	9.6

The table 11 illustrates examples of images corresponding to confidence scores of 1.8, 6, and 9.6, providing a visual explanation of what each score signifies in terms of image quality.

11.3. Key Parameters Selection

Since the crowdsourced data is collected during the daytime under clear weather conditions, i.e., without rain, snow, fog, or sand, the primary quality check parameters used in the evaluation are:

- Blur (Daytime)
- Illumination
- Lane Line Degradation
- Presence of Objects on the Lanes
- Visibility of Lane Lines

12. Detailed Discussion on Results of Real Crowdsourced Vehicle Collected Data

12.1. Analysis of Confidence Scores for Local Maps

The MLLM-driven confidence scoring model plays a crucial role in filtering out low-quality image sequences, directly impacting the accuracy of HD maps. Table 5 provides insights into the quality of image sequences captured across different link areas in multiple local maps, as determined by the average confidence score.

- **High Confidence Scores for Local Maps 1–3:** The first three local maps in each link area consistently exhibit high confidence scores, ranging between an average of 7.62 and 8.30. These results suggest that images in these maps were captured under favorable conditions, where key parameters such as lane line visibility, illumination, and the absence of blur were optimal. This indicates clear lane markings and suitable weather and road conditions, making this data highly reliable for HD map updates.
  - In Link Area 6, Local Map 1 has a high score of 8.80, Local Map 2 scores 8.46, and Local Map 3 scores 7.82, indicating optimal data quality.
  - Similarly, Link Area 73 shows a strong confidence score of 8.57 for Local Map 1, followed by 7.82 and 7.75 for Local Maps 2 and 3, confirming good lane visibility and image clarity.
- **Confidence Scores in Local Maps 4 and 5:** A decline in confidence scores for Local Maps 4 and 5 across most link areas suggests challenges in the data collection process. The scores for these maps range between 5.38 and 6.87, indicating image quality degradation, likely due to high illumination, glare, or lane degradation.
  - In Link Area 6, the confidence score drops from 7.82 in Local Map 3 to 5.38 in Local Map 5. This reduction could be attributed to excessive glare, blurred imagery, or obscured lane lines, making the data less reliable for HD map updates.
  - Similarly, Link Area 67 experiences a decline from 7.25 in Local Map 3 to 5.96 in Local Map 5, suggesting deteriorating conditions such as blurred images, poor lane visibility, or traffic obstructions.
- **Filtering Data Using Confidence Scores:** The confidence scores generated by the Data Cleansing Model help identify local maps with valid and high-quality image sequences suitable for HD map updates. By selecting an appropriate threshold, unreliable maps can be filtered out. For instance, setting a threshold confidence score of 7.0

ensures that any local map below this value is excluded from the HD map update process.

This approach demonstrates that:

– Only the highest-quality data is used for map updates, increasing the overall accuracy and reliability of the HD map.
– Maps with confidence scores above the threshold are captured under favorable conditions, ensuring good lane visibility, minimal blur, and optimal illumination.

12.2. Integration of Detected Changes into the HD Map

The next phase of the HD map update process focuses on incorporating detected changes into the existing map for several link areas, as illustrated in Figures 15, 16, and 17. Three link areas were selected where specific modifications were identified and then fused with local maps containing high-confidence image sequences to generate an updated map. These changes primarily involve shifting lane lines, which includes the removal of outdated lane markings and the addition of new ones.

The ICP-based association ensures that shifted lanes are accurately aligned with the original map, while DBSCAN handles fusion by incorporating new lanes and removing obsolete ones. Such updates are crucial for maintaining accurate and safe navigation, particularly in dynamic environments where road conditions frequently change.

In Figures 15, 16, and 17, the left side represents the Changed/Modified New Map, while the right side displays the Fused Map, highlighting the differences. The yellow color in the modified map signifies lane shifts, while green indicates newly added lanes. In the fused map, green represents lane shifts, and blue represents newly added lanes. The selected link areas, each exhibiting unique structural characteristics and modifications, emphasize the importance of these updates.

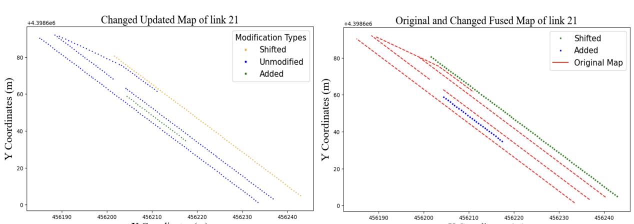


Figure 15. Visualisation of Changed New Map and Fused Map in Link Area 21.

The fused map integrates the high-confidence local map with detected changes, ensuring an accurate representation of the current road layout.

- **Link Area 21:** As shown in Figure 15, Link Area 21 exhibits a pattern of shifted and newly added lanes. The detected changes (left) indicate that several lane lines in the upper-left quadrant required shifting, while new lanes were added toward the center. The fused updated map (right) demonstrates the successful integration of these modifications. Since Link Area 21 includes high-traffic areas, confidence score-based data selection played a crucial role in selecting optimal sequences, minimizing discrepancies with the ground truth map, and ensuring precise lane alignment.

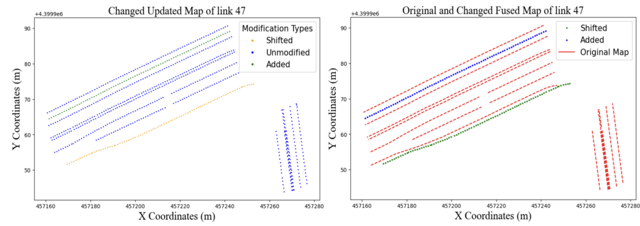


Figure 16. Visualisation of Changed New Map and Fused Map in Link Area 47.

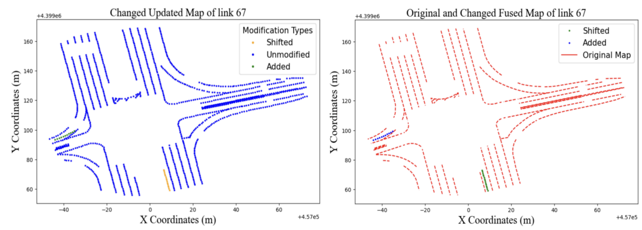


Figure 17. Visualisation of Changed New Map and Fused Map in Link Area 67.

- **Link Area 47:** As shown in Figure 16, Link Area 47 demonstrates more extensive modifications. The detected changes (left) indicate multiple shifted lanes, particularly in the bottom-right region. The fused map (right) illustrates how these updates were integrated, with new lane lines reflecting the current road layout. Accurate local map sequences were essential in ensuring that the modifications aligned correctly with the updated lane configurations.
- **Link Area 67:** As shown in Figure 17, Link Area 67 has a complex structure with multiple intersections and branching lane lines. The detected changes (left) reveal multiple shifted lane lines, particularly near intersections, which required precise alignment. The fused map (right) integrates these changes smoothly, accurately reflecting the new configurations and enhancing routing information for navigation. In such complex environments, confidence score-based map selection is vital to capturing intricate details like lane transitions and merges with high precision.

12.3. Physical Meaning of the terms Seq1, Seq3 and Seq5

In the evaluation presented in Table 6, a sequence Seqk refers to the fusion of top k local maps ranked based on their average confidence scores while performing map update:

- **Baseline:** Updates are performed by fusing all available local maps, irrespective of their confidence scores. This approach includes both high and low-quality data, maximizing the number of data points but potentially reducing overall map reliability.
- **Seq1:** Updates are made using only the local map with the highest confidence score, ensuring that only the most reliable data is used. This sequence involves the minimum number of data points but maintains the highest data quality.
- **Seq3:** The map is updated by fusing top 3 local maps i.e. first using the local map with the highest confidence score, then incorporating fusion of the second- and third-best ranked maps respectively. This sequence increases the number of data points while maintaining relatively high data quality.
- **Seq5:** The map is updated by fusing top 5 local maps i.e., from top 1 till top 5 highest scoring local maps. While this increases the data points further, the inclusion of lower-confidence maps introduces lower-quality data into the update process.
- **MiniGPT:** Updates are performed by retaining all images and their corresponding lane line data points with confidence scores of 7 or higher. This ensures that only data from local maps containing images above the confidence threshold (considered to represent good quality) is used. The confidence scores are calculated by MiniGPT-v2, which follows predefined rules emphasizing image clarity.

12.4. Optimal Sequence for HD Map Update

Based on the evaluation results in Table 6, 7 and Figure 6, Seq3 is identified as the most optimal sequence for updating the HD map. It balances data quality and quantity, ensuring that the updated map remains accurate while incorporating sufficient data points to handle complex road configurations. This sequence offers several advantages:

- **Higher Data Quality:** Seq3 maintains a high average confidence score of 7.6, which is close to the best possible score of 8.3.
- **Sufficient Data Points:** By incorporating three local maps, Seq3 ensures that enough data points are included to accurately model the map without introducing excessive noise.
- **Low Error:** The average mean error across all link areas for Seq3 is 0.28 meters, significantly lower than the baseline as well as other methods' average mean error of 0.37 meters, demonstrating that the system maintains accuracy even with increased data points. Furthermore, the error for Seq3 falls well below the minimum accuracy requirement for lane lines in HD maps, which is less than equal to 0.32 meters, as established by Křehlík et al. (2023) [49]

12.5. Trade-off Between Confidence Score and Data Points

One of the key insights from the evaluation results is the trade-off between confidence scores and the number of data points. Confidence scores indicate the quality of an image sequence, where higher scores correspond to optimal conditions such as clear lane visibility, minimal blur, and proper illumination. However, increasing the number of data points often requires incorporating local maps with lower confidence scores, introducing noise and reducing overall accuracy.

This trade-off is evident in the performance of different sequences:

- **Seq1:** Utilizes only the local map with the highest confidence score, ensuring minimal error (0.30 meters on average). Although it limits the number of data points, the high data quality results in accurate map updates.
- **Seq3:** Expands on Seq1 by incorporating the second and third best-scoring local maps. This increases the number of data points while maintaining a low error rate of 0.28 meters, making Seq3 the most optimal configuration. It effectively balances data quality with sufficient data points for accurate map updates.
- **Seq5:** Adds lower-confidence local maps (fourth and fifth), leading to an increase in error to 0.34 meters. While it introduces more data points, the inclusion of lower-quality maps degrades overall accuracy.

The trade-off demonstrates that while adding more data points can improve map coverage, incorporating lower-confidence maps introduces errors. Managing this balance is crucial for maintaining both accuracy and coverage in HD map updates.

12.6. Optimal Confidence Score Threshold for HD Map Updates

Another critical observation from the results is the importance of setting an appropriate confidence score threshold for selecting data in HD map updates. As shown in Table 5, the first three local maps in each link area have high confidence scores ranging from 7.6 to 8.8, correlating with lower error rates in Seq1 and Seq3. In contrast, local maps 4 and 5, with confidence scores between 5.8 and 6.6, introduce greater errors in Seq5.

Based on these findings, it is recommended to set a confidence score threshold of 7.0 or higher for HD map updates. This threshold ensures that only high-quality data is used, reducing the likelihood of introducing errors due to lower-quality data. A threshold of 7.0 effectively balances data quality with the number of data points, as evidenced by Seq3, which achieves an optimal configuration.

12.7. Reliability of the MLLM-Driven Confidence Score-Based HD Map Update

The results demonstrate that the confidence score-based HD map update system is highly reliable. The system consistently outperforms the baseline in all cases, with Seq1 and Seq3 achieving significantly lower errors. Even Seq5, despite incorporating lower-confidence data, performs better than the baseline, proving that the confidence score-based approach effectively filters poor-quality data while utilizing high-quality inputs.

The robustness of this approach lies in its ability to minimize noise by prioritizing high-confidence local maps, ensuring highly accurate updated maps. The error values obtained from Seq1 and Seq3 meet the minimum accuracy requirements for HD maps, further affirming the reliability of this approach for real-world applications in autonomous vehicle navigation.

The confidence score-based HD map update system effectively maintains the accuracy and reliability of HD maps. The trade-off between confidence score and the number of data points is a crucial factor in the update process, and the results indicate that Seq3 provides the best balance between these two elements. By using the top three local maps with the highest confidence scores, Seq3 ensures both high data quality and sufficient data points, leading to accurate map updates.

Furthermore, the analysis supports setting a confidence score threshold of 7.0 to ensure that only high-quality data contributes to map updates. This threshold minimizes the introduction of errors while maintaining comprehensive map coverage. The results confirm that Seq3 provides the optimal configuration for HD map updates, achieving a mean error significantly lower than the baseline and meeting the accuracy requirements for autonomous navigation systems.

The confidence score-based approach not only enhances accuracy but also ensures the reliability of HD maps, making it an ideal solution for large-scale HD map updates. This system strengthens the ability of autonomous vehicles to navigate complex road environments with precision and safety.

13. Supplementary Conclusion

The evaluation of the confidence score-based HD map update system has provided several key insights into optimizing map accuracy using quality-assessed data. The experiments demonstrated that the proposed approach consistently outperforms the baseline, showing a significant reduction in mean errors across different sequences. Notably, Seq3 was identified as the most optimal configuration, achieving a mean error of 0.28 meters, compared to the baseline of 0.37 meters. By using the top three local maps based on their confidence scores, Seq3 managed to strike a balance between data quality and quantity, ensuring comprehensive coverage while maintaining high precision.

These findings confirm the efficacy of confidence score-driven methodology for large-scale HD map update systems, supporting safer and more precise autonomous vehicle navigation. This framework sets a strong foundation for improving autonomous vehicle navigation through more accurate and adaptive map updates.