International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 12-No.1 (2026) pp. 506-514
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Efficient Ingestion, Labeling, and Storage of LiDAR, Radar, and Camera Datasets: Optimizing Storage Formats for Autonomous Vehicle Development

Yogesh Pugazhendhi Duraisamy Rajamani*

Independent Researcher, USA
* **Corresponding Author Email:** yogeshdurairajamani@gmail.com - **ORCID:** 0000-0002-0047-885X

Article Info:

DOI: 10.22399/ijcesen.4839
Received : 28 November 2025
Revised : 26 January 2026
Accepted : 28 January 2026

Keywords

Autonomous Vehicle Perception,
Multi-Modal Sensor Fusion,
Lidar Point Cloud Processing,
Dataset Management Infrastructure,
Storage Optimization

Abstract:

The development of autonomous vehicle technologies has placed new standards of managing the multimodal sensor data that includes LiDAR point clouds, radar measurements, and camera images. The contemporary autonomous platforms produce large volumes of non-uniform perception data that need advanced infrastructure to ingest, label, and store in the long term. Combining different modalities of sensing creates significant technical issues throughout the data lifetime, both during the initial capture and archival preservation. Infrastructure Edge preprocessing hardware will need to synchronize heterogeneous sensor streams with less than millisecond accuracy and apply real-time compression to achieve a lower bandwidth in transmission. Automated labeling workflows based on pre-trained perception models can help reduce the manual annotation load by a significant factor, and label quality is likely to be high enough to be used in training. Storage format optimization trades off conflicting demands such as compression performance, random access performance, and compatibility with the distributed processing model. The use of cloud-scale deployment allows managing the petabyte-scale volumes of data and optimizing the costs of infrastructure based on the tiered storage approaches that match the storage performance properties with the access patterns. Benchmark datasets have played a vital role in achieving advances in autonomous vehicle perception, and have allowed the performance to be systematically compared and the outstanding challenges to be reflected. Direct neural network processing of point cloud data has permitted significant progress in performance in detection and segmentation. Further progress will be realised through further growing multi-modal datasets with multi-modes of operation and persistent advances in algorithms that fuse sensor data and build scene understanding.

1. Introduction

The development of autonomous vehicle technologies has radically changed the research and development of transportation, introducing new opportunities in the management of large amounts of multimodal sensor data. Modern autonomous driving systems are based on advanced perception pipelines, combining various sensing modalities into the development of complex environmental representations. The exponential increase in the data rate created by operational test fleets has led to the need to develop strong infrastructure that can provide sensor data at scale levels that had not been experienced before in the automotive engineering field. Since their development, modern autonomous platforms use multiple types of sensors at the same time, all of which record different elements of the driving environment using varying temporal resolutions, spatial attributes, and semantic information content. The diversity of multi-sensor configurations poses significant technical problems throughout the life cycle of the data, including capture and long-term archival storage. Perception sensors produce continuous streams of high-dimensional measurements, which need to be correlated correctly, calibrated, and efficient to facilitate real-time operation decision-making processes and offline model development processes. Close attention to the architectural design is needed with the integration of LiDAR point clouds, which give detailed geometric features, radar data which give strong velocity estimates even in adverse weather conditions, and camera data which give rich semantic features. The paper has detailed approaches to effective multi-modal sensor data management, which include ingestion strategies tailored to edge computing requirements, automated labeling workflows that reduce the high burden of manual annotation, and storage optimization strategies that balance access performance with infrastructure cost and data fidelity needs of machine learning applications.

2. Characteristics and Properties of Multi-Modal Sensor Data

2.1 LiDAR Point Cloud Characteristics and Segmentation

LDR technology has become one of the foundation sensing modalities of autonomous vehicle perception systems, offering dense 3D measurements of the surrounding environment by laser-based distance measurement. LiDAR point clouds pose special computational problems because of the sporadic spatial sampling, as well as large numbers of points that are produced in real-time as the vehicle moves. Studies on fast segmentation of the LiDAR images have indicated that ground plane extraction is a very important preprocessing stage, since it allows isolated strips of the drivable surface area and obstacle points that need further processing [1]. The autonomous vehicle application segmentation paradigm focuses on the real-time processing power, and hence, algorithmic solutions must be able to run within a strict notion of computational constraints and must also be able to provide a high-quality performance under a variety of environmental factors. Ground segmentation algorithms should be capable of supporting the geometries of the terrain, from flat highway surfaces to complicated urban terrain with slope, curbs and elevation variations. LiDAR point cloud density also depends greatly with range in the sense that algorithms need to maintain the same level of performance throughout the entire sensing volume between direct vehicle proximity and the maximum possible range. Mechanical scanning LiDAR systems have temporal properties causing motion distortion effects, in that the vehicle platform does not stand still during each scan cycle and must be compensated using motion compensation methods that take into consideration the ego-vehicle motion. The reflectivity characteristics of LiDAR sensors give useful extra information with respect to pure geometry, which allows a material-based classification that separates road surfaces, vegetation, vehicles, and other categories of objects.

Semantic interpretation of LiDAR sequences is not limited to the analysis of a single frame, but instead it can be applied over time, to successive measurements. The creation of large-scale datasets for semantic scene understanding has now made it possible to systematically evaluate the perception algorithms on long driving sequences [2]. The datasets contain per-sample semantic labeling of dense point clouds of a wide variety of object classes present on the roads in real-world driving. LiDAR sequence annotation should be performed with close consideration of temporal coherence, whereby the labels of the objects should be consistent as the vehicle traverses the space and as moving objects themselves display their dynamic motions. The capability of distinguishing between subtle geometric and intensity patterns that distinguish between similar object categories is essential to semantic segmentation performance. The existence of temporally long sequences makes it possible to create algorithms that take advantage of motion information and temporal consistency constraints to achieve better segmentation results than are possible with single-frame analysis. Studies have shown that multi-scale analysis methods are useful in semantic understanding because they capture local geometry on a fine scale and global patterns on a broad scale. Combining semantic segmentation outputs with object detection and tracking pipelines produces full scene understanding functions, which are needed in safe autonomous navigation.

2.2 Multi-Modal Dataset Characteristics

The development of large-scale autonomous driving data has been motivated by the realization that sound perception systems need training data that is representative of the full containment of situations of operation. Multi-modal data sets are complete sets of measurements of several sensor types, where the measurements are synchronized, and the application of fusion algorithms is possible to utilize complementary sensing properties [3]. A2D2 data set is an illustrative example of the contemporary systems of autonomous driving data collection, where the large sensor packages are installed on the test vehicles that are driving across different geographic areas and climate conditions. The design of datasets is considered to include sensor choice, mounting policies, calibration policies, and synchronization infrastructure needed to keep the different data streams in time synchronization. The geographic diversity of inclusive datasets guarantees that the autonomous systems handle with consistent reliability different road infrastructure, traffic flow, and environmental conditions. Weather variability is a special problem because the performance of the perception tends to suffer in unfavorable conditions such as rain, fog, and low-light situations, which are frequent in real-world operations.

The factor of scalability has grown in prominence as autonomous vehicle development projects grow in maturity and scale of running test fleets [4]. The Waymo Open Dataset shows the magnitude possible with the help of long-term data collection activities, including the multiple hours of driving in various places and environments. The size of datasets directly affects the statistical coverage of the situation of rare but safety-sensitive parameters that should be addressed properly in spite of their relative rarity. Large-scale datasets also pose significant logistical difficulties when they are going to be annotated, necessitating distributed annotation forces and advanced quality control processes to ensure that labels are consistent. The distribution of datasets needs to trade off the completeness of the data with the practical issues such as storage space, distribution infrastructure, and documentation necessary to facilitate meaningful use of the research community. Normalization of data format and testing guidelines enables meaningful comparison of the performance of various algorithms in a systematic way to boost development by setting standardized benchmarks.

3. Multi-Sensor Ingestion Pipeline Architecture

3.1 Synchronized Multi-Modal Data Capture

The time synchronization of the sensors used in the construction of the effective multi-modal perception datasets must be treated with care, such that the measurements of the various modalities reflect the same environmental conditions. The nuScenes dataset was the first to annotate synchronized sensor suites, and it set methods for temporal alignment between cameras, LiDAR, or radar sensors [5]. The accuracy of synchronization has a direct effect on sensor fusion algorithms, whereby the presence of a time error leads to systematic errors in a sensor fusion algorithm that worsen the performance of that algorithm. Hardware-based synchronization methods offer better performance on time when compared to software-based methods, and such mechanisms apply common clock signals in every sensor module by means of specific triggering infrastructure. The convention of the coordinate frame to use in sensor fusion is a matter of serious concern because transformation conventions may affect subsequent algorithm development and compatibility between datasets. Calibration processes allow the establishment of the accurate geometric relationship between sensor reference frames, which allows measurement to be projected into common coordinate systems where the fusion operations take place.

The ingestion pipeline should support different rates of updates between different sensor types, and LiDAR systems usually run with constant scan frequencies when compared to camera frame rates, which can be variable depending on exposure needs, and radar systems provide detections asynchronously. Buffering schemes allow synchronization of asynchronous sensor streams; they gather measurements inside temporal windows and match them by temporal proximity to one another. The delivery of synchronized sensor data needs format planning capable of conserving time-based associations on the one hand and thus providing effective storage and recovery on the other. Information that comes with sensor measurements is metadata, such as latent parameters of the sensor, sensor condition, and a description of the environmental condition, which gives meaning to future analysis.

3.1.1 Detailed Data Ingestion Process Summary

The multi-sensor data ingestion process operates through a sophisticated multi-stage pipeline designed to handle heterogeneous sensor streams efficiently while maintaining temporal coherence and data integrity. The complete ingestion workflow encompasses five primary stages: sensor acquisition, temporal synchronization, edge preprocessing, data packaging, and cloud transmission.

Sensor Acquisition Stage: The initial stage involves simultaneous data capture from multiple sensor modalities mounted on autonomous vehicle platforms. LiDAR sensors generate dense point clouds at fixed scanning frequencies, typically producing data streams at sustained rates. Radar systems provide sparse object detections asynchronously with varying update frequencies based on target detection events. Camera arrays capture high-resolution imagery with frame rates that adapt dynamically based on exposure requirements and ambient lighting conditions. GPS and inertial measurement units provide precise positioning and timing references that serve as the temporal foundation for multi-sensor synchronization. [4]

Temporal Synchronization Stage: Hardware-based synchronization mechanisms distribute precision clock signals to all sensor modules through dedicated triggering infrastructure, achieving sub-microsecond alignment accuracy. The synchronization system employs Precision Time Protocol or GPS pulse-per-second signals as master clock references. Buffering schemes collect measurements from asynchronous sensor streams within temporal windows, associating data based on timestamp proximity. The temporal alignment process accounts for varying sensor latencies and ensures that fused measurements correspond to consistent environmental states despite different acquisition rates across modalities[5].

Edge Preprocessing Stage: On-vehicle edge computing platforms perform real-time preprocessing to reduce downstream bandwidth requirements and prepare data for storage. Calibration modules apply intrinsic and extrinsic transformation parameters, converting sensor measurements from individual reference frames into unified vehicle coordinate systems. Compression algorithms optimized for different data types reduce storage footprints while maintaining perceptual quality, with LiDAR point clouds processed through geometric compression and camera streams encoded using video codecs. Filtering operations perform selective downsampling and event-based capture, retaining high-value data segments while discarding redundant information. Motion compensation algorithms correct for ego-vehicle movement during sensor acquisition periods, particularly critical for mechanical scanning LiDAR systems. [6]

Data Packaging Stage: Preprocessed sensor streams are encapsulated into standardized container formats that preserve temporal relationships and embed comprehensive metadata. Modern packaging approaches utilize formats such as ROS2 MCAP or Protocol Buffers that support efficient random access and chunk-based organization. Metadata embedded within packages includes calibration parameters, sensor configuration descriptors, vehicle state information, environmental condition tags, and scenario-specific annotations. The packaging process organizes multi-modal data into coherent units suitable for subsequent storage, retrieval, and analysis workflows.

Cloud Transmission and Storage Stage: Packaged sensor data transmits from vehicle platforms to cloud infrastructure through multiple communication channels optimized for different data priorities and network conditions. Real-time telemetry streams utilize lightweight protocols for continuous transmission of critical information over cellular networks. Bulk sensor data uploads occur opportunistically when vehicles connect to high-bandwidth WiFi infrastructure, typically at depot charging locations. Cloud ingestion layers distribute incoming data across storage tiers based on access patterns, with active training datasets stored in high-performance hot storage, labeled datasets maintained in warm storage, and archival data migrated to cost-optimized cold storage. Distributed processing frameworks enable parallel analysis of ingested data across compute clusters, while metadata indexing systems provide rapid query capabilities across petabyte-scale datasets. [5]

3.2 Real-Time Processing and Detection

Real-time perception pipelines are based on efficient object detection algorithms that are run on incoming sensor streams with latency constraints determined by control system requirements. Bird-eye view representations have turned out to be useful intermediates to LiDAR-based detection, casting three-dimensional point clouds on two-dimensional planes, where standard convolutional neural networks can be effectively used [6]. PIXOR architecture shows that real-time detection can be performed with well-constructed network architectures that can take advantage of the spatial structure available in organized point cloud images. Single-stage detectors do not use region proposal methods like those in two-stage detectors and instead regress detection bounding boxes to feature maps in one forward network pass. The estimation of the orientation of objects detected poses specific difficulties as the heading angle is a continuous circular variable, which needs special loss functions construction.

Autonomous vehicles have an edge computing platform with limited computational resources relative to the cloud computing infrastructure, which requires the design of algorithms that optimize the accuracy of detections and the efficiency of the computation. Quantization and pruning are model compression methods that allow the execution of advanced neural networks on resource-limited edge devices. Backbone architecture choice influences inference latency and detection accuracy, and there is current research on computing the best trade-offs between the two in automotive systems. Learning methods of multi-tasking can achieve effective sharing of computational resources among similar perception problems, sharing the cost of feature extraction among detection, segmentation, and classification tasks.

4. Automated Labeling and Annotation Workflows

4.1 Annotation Methodologies and Tooling

Production of quality training data involves the use of advanced annotation tools that allow the effective labelling of multimodal sensor suites of complex three-dimensional scenes. Annotation interfaces should enable easy manipulation of bounding boxes in 3-D space and display correlated views of multiple sensor modalities to make sure that they are consistent. Three-dimensional scene annotation is more complex than the two-dimensional image labelling used in traditional labelling tasks because annotators are required to make three-dimensional reasoning about object extents in a three-dimensional space, taking into consideration occlusions and partial visibility. Quality control measures guarantee the consistency of annotation in large datasets, which are both checked automatically and manually. Training annotation workforce members is associated with a significant investment because three-dimensional scene understanding and the efficient use of the tools are area-specific skills. The rules of annotation should be cautious about how to deal with ambiguous ones, as well as the delimitation of the boundary of the object extent, and the labeling of the object that is partially visible.Standardized annotation protocols help the annotation teams to be consistent and the labels produced by other annotation workflows to be compared. Temporal annotation of dynamic scenes adds even more complexity, as an object identity needs to be tracked over consecutive frames, even though the viewpoint and the presence of partial occlusions change. Of special concern is the annotation of rare scenarios, which are statistically rare but whose correct treatment is important to the safety of the system. The annotation tools are becoming more and more 3D visualizations, such that one can quickly gain an intuitive grasp of the intricate geometry of an annotated scene.

4.2 Neural Network Architectures for Point Cloud Processing

The deep learning architecture of working directly with unordered point sets has contributed to revolutionizing the application of deep learning to point cloud data. The techniques used in the early works of three-dimensional shape recognition were multi-view rendering methods, creating two-dimensional projections that were inputtable into standard convolutional networks [7]. The multi-view paradigm also shows that three-dimensional knowledge can be obtained by aggregation of many two-dimensional views, but this paradigm has the disadvantage of adding computational complexity through rendering operations, and the paradigm can discard information that exists in the original three-dimensional representation. Strategies of viewpoint selection influence the extent of coverage of shape produced by multi-view strategies, with well-selected sets of viewpoints giving greater coverage to shape characterization. This combination of characteristics derived across different views must be aggregated using information combination schemes that are invariant to the order of the viewpoints.

The paradigm shift in the pointnet architecture was that point clouds could be handled by neural networks without converting them into volumetric or multi-view representations [8]. The permutation invariance obtained by a symmetric aggregation operation allows a network to process point clouds in any order, and this is a particularly important feature because point sets are unordered. Hierarchical extension of point cloud processing: The networks have the ability to record both multi-scale geometric features due to progressive spatial grouping actions. Local geometric features, which are captured at fine scales, are added to create a higher-level semantic knowledge by hierarchical abstraction. The successful operation of direct point cloud processing has provoked a thorough research on architectural variations and extensions to particular perception tasks.

5. Storage Format Optimization for Multi-Modal Datasets

5.1 Semantic Scene Understanding Datasets

Large-scale datasets, which have to be organized and stored in a manner that facilitates semantic scene understanding, should take into account access patterns and query specifications. The rich annotation of urban scenes can be used to perform semantic analysis of the built environment at a fine level to address autonomous navigation as well as city planning [9]. The Cityscapes dataset is an example of a good semantic annotation of urban driving scenes, which gives pixel-wise labels of a wide variety of object types found in the urban setting. The granularity of the annotation allows semantic segmentation models to be trained and to be able to produce detailed parsing of the scene. Stereo imagery offers geometric information to support semantic information, and through this, algorithms are able to reason about the structure of the scene and semantic content. The variety of scenes that were captured across various cities makes the coverage of various architectural designs, road geometry, and urban formations that are experienced in practice in real-life operations.

The organization strategies of datasets influence the efficiency of loading data into a model, and a well-thought-out directory structure and file formats minimise I/O bottlenecks. The division of datasets into training, validation, and test subsets would be done by paying attention to the diversity of scenes in order to make sure that they are represented by the partitions. Metadata organization allows efficient subsetting and querying of data sets by their characteristics in a scene, which are useful in analyzing algorithm behavior on a specific type of scenario. The recorded characteristics of the datasets, the procedure of annotation, and known limitations ensure that the dataset is used appropriately by the research community.

5.2 Benchmark Dataset Infrastructure

Setting up universal standards has played a crucial role in the advancement of autonomous vehicle perception studies. The KITTI vision benchmark suite was the first to systematically test perception algorithms on a variety of tasks such as object detection, object tracking, and depth estimation [10]. The design of the benchmark should take into account evaluation metrics that reflect areas of performance that are applicable in the actual deployment of the system. Public leaderboards allow the clear comparison of the performance of various algorithmic strategies, which leads to the further evolution of enhanced strategies. Assessment plans should describe exactly the circumstances in which algorithms should be utilized, such as what inputs are allowed, what preprocessing is allowed, and whether or not temporal context is available. The difficulty ranges of benchmarks allow one to examine the performance of algorithms (especially the existing algorithms) more finely, separating between the easy cases, where most algorithms can perform, and the challenging ones, where the capabilities of the specific algorithm are evident.

The data division should be in such a way that there is no leakage of information in training to test data sets, so that the performance appraisals made are accurate. The presence of test set annotations with the organizers of the benchmark hinders overfitting to test data by submitting the same model to the benchmark multiple times in order to optimize its score. Continued relevance with improved algorithms and new methods leads to benchmark longevity, and initial design and updates at times with the field to changes. Public training and validation data facilitate training research and ensure the integrity of evaluations because of standardized development platforms, whereas test annotations are not shared.

6. Cloud-Scale Deployment and Best Practices

6.1 Distributed Processing Infrastructure

A distributed computing infrastructure is necessary to process large data volumes of autonomous vehicles, since autonomous vehicles need large data volumes to operate. The cloud-based processing frameworks allow data processing pipelines to be executed in parallel by a large number of compute nodes with throughputs that would not be possible on individual machines. Scalable processing pipeline design must pay close attention to data partitioning techniques that can be used to provide independent processing of data subsets and reduce the overhead of inter-node communication. Cloud-scale management of data is based on a distributed storage system that offers aggregate bandwidth capable of supporting numerous parallel processing tasks. The nature of the workloads determines the kind of distributed computing framework to use as a batch-oriented framework is capable of performing offline processing and a streaming framework is capable of performing a near-real-time analysis of data.

Resource management systems coordinate the scheduling of the consumption of computational resources among competing workloads to achieve the efficient use of available infrastructure. Container-based deployment strategies allow the reproducible execution environments and ease the process of moving processing workloads across cloud providers. Cost optimization mechanisms trade the computational throughput on infrastructure with the infrastructural costs, using spot instances to support fault-tolerant batch workloads and using guaranteed capacity to support latency-sensitive workloads. Tracking of infrastructure follows the processing throughput and determines bottlenecks which allows ongoing optimization of pipeline efficiency.

6.2 Data Lifecycle Management

Data lifecycle management involves the entire process of data from the time it is first captured until its final archival or deletion. The retention policies should strike a compromise between the worth of the old data and the expenses of storing the data in the long run, based on the cost of maintaining the infrastructure and the possible future use. Tiered storage strategies that can optimize cost without significant access latency to active datasets are made possible by classifying data by business value and access frequency. Data is automatically moved between storage levels according to age and access patterns, minimising the chance of manual controls.

The data governance frameworks will take care of the proper management of information that may be sensitive, and that may have been inadvertently taken in the process of data collection. The access control systems ensure that only authorized staff can access the data to avoid unauthorized leakage of proprietary data or information of privacy significance. Audit logging follows the access pattern of data, which helps to execute compliance checks and security incident investigations. Datasets and processing pipelines versioning guarantee the reproducibility of the research results and allow reversion to the earlier versions in case something wrong has been detected in the latest releases.

7. Challenges and Future Directions

The further development of autonomous vehicle perception systems is fraught with a lot of technical complexities, necessitating the consistent research and development of the system. This has raised an open challenge because the generalization of the perception algorithms to different functional areas is not always desirable. In most cases, a model that is trained on data in certain geographic areas or weather conditions may show poor performance when applied to new environments. Domain adaptation methods aim to address the discrepancies in performance between the training and deployment settings, utilizing unlabeled target domain data to refine the source domain-trained models. New sensor technologies, such as solid-state LiDAR and high-resolution imaging radar, present new data properties unrepresented by existing datasets, with the need for additional annotation and algorithm modification.

There is a need to carefully design the interface between learned perception modules and classical robotics structures in order to ensure that outputs of the neural network deliver the semantic and geometric information that is needed later by downstream planning and control systems. This verification and validation of learned perception systems are difficult problems with special consideration to the traditional software because the neural networks cannot be fully tested on all the potential input problems, and to learn neural networks, which are verifiable, is computationally infeasible. Testing methods with simulation provide synthetic sensor data reflecting varied situations, allowing it to exhaustively cover operational states, which cannot be done during physical testing. Changes between simulated sensor data and real sensor data have been an ongoing problem with perception algorithms trained on non-real data in the literature, with poorer performance on real-world measurements of sensors being reported because of small variations in sensor properties and rendering of the environment.

The opportunities and challenges of deployments of autonomous vehicles are in the development of online learning systems that are constantly enhanced depending on the operational data. Constant learning methods should not be too plastic or too stable to maintain performance in already learned situations. Distribution shifts, which signal that deployed models find themselves in scenarios the training distribution does not cover, allow the proactive updating of the models before the failure can take place, which is potentially safety-critical. Privacy-sensitive learning algorithms make it possible to apply model enhancement when pieced together with a distributed fleet of vehicles without necessarily having to collect potentially sensitive raw sensor measurements centrally.

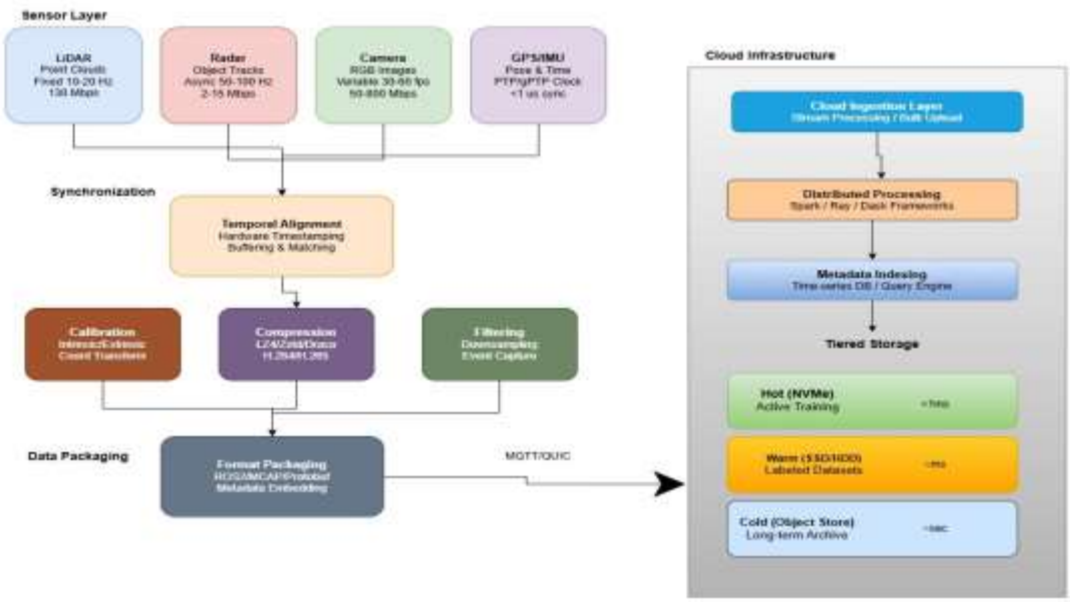

***Figure 1**: Enhanced Multi-Sensor Ingestion Pipeline Architecture [4, 5, 6]*

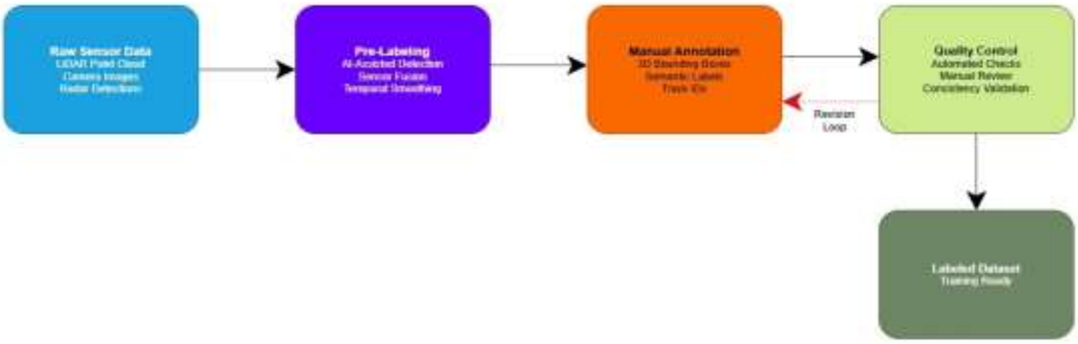

Figure 2: *Annotation Workflow Pipeline [5, 6]*

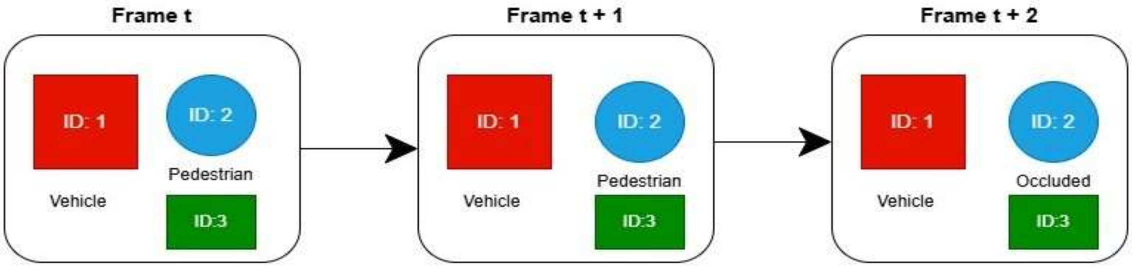

Figure 3: *Temporal Tracking Across Frames [5, 6, 7, 8]*

Table 1: *Multi-Modal Sensor Characteristics Comparison [2, 3]*

Sensor Type	Data Structure	Primary Information	Temporal Resolution	Processing Challenge
LiDAR	Point Clouds	3D Geometry, Intensity	Fixed Scan Rate	Motion Distortion
Radar	Sparse Detections	Range, Velocity, RCS	Asynchronous	Noise Filtering
Camera	Dense Imagery	Semantic Content	Variable Frame Rate	Compression Quality

Table 2: *Storage Format Optimization Comparison [8, 9, 10]*

Format Type	Compression Ratio	Access Pattern	Best Use Case
MCAP	Moderate	Random Seek	Multi-sensor Fusion
HDF5	High	Hierarchical Query	Scientific Computing
Parquet	Very High	Columnar Analysis	Metadata Indexing
Binary Custom	Configurable	Sequential Streaming	GPU Training Pipeline

Table 3: *Cloud Deployment Storage Tiers [8, 9, 10]*

Tier Level	Access Latency	Cost Factor	Typical Content
Hot (NVMe)	Sub-millisecond	Premium	Active Training Data
Warm (SSD/HDD)	Milliseconds	Moderate	Labeled Datasets
Cold (Object Storage)	Seconds	Minimal	Long-term Archive
Glacier	Minutes-Hours	Ultra-low	Compliance Records

8. Conclusions

Efficient handling of multi-modal sensor data forms a key facilitator of autonomous vehicle solutions, and it demands all-encompassing solutions on the data capture, data annotation, data storage, and data analysis aspects. Sturdy edge processing hardware with the ability to coordinate heterogeneous sensor streams and also implement compression and filter mechanisms saves a significant amount of transmission bandwidth. Automated labeling processes based on pre-trained perception models do not require much manual annotation, but label quality is adequate in those applications where the model is trained. Storage format optimization has been solving conflicting needs such as compression efficiency, random access performance, and compatibility with distributed processing frameworks needed in large-scale machine learning operations. Scalability through cloud-based implementation of dataset management infrastructure allows for maintaining petabytes of data and reducing costs with tiered storage policies. Large benchmark datasets have facilitated advancements in autonomous vehicle perception, facilitating comparative performance evaluation on a systematic basis and identifying issues that need additional focus. The semantic interpretation of the complex urban environment necessitates the combination of the geometric data of the complex urban environment collected by means of LiDAR sensors with the semantic data of the complex urban environment collected by means of camera images and radar systems. The ability to run point cloud data directly on specialized neural network architectures has made it possible to improve the performance of detection and segmentation tasks and also lower the preprocessing overheads. It will continue to be expanded in the future with multi-modal datasets of a variety of operational conditions, further algorithm innovation in sensor fusion, and the implementation of more capable edge computing infrastructure to support advanced on-vehicle processing capabilities.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.

- **Author contributions:** The authors declare that they have equal right on this paper.

- **Funding information:** The authors declare that there is no funding to be acknowledged.

- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] Dimitris Zermas et al., "Fast Segmentation of 3D Point Clouds: A Paradigm on LiDAR Data for Autonomous Vehicle Applications," ResearchGate, 2017. [Online]. Available: https://www.researchgate.net/publication/318325507_Fast_Segmentation_of_3D_Point_Clouds_A_Paradigm_on_LiDAR_Data_for_Autonomous_Vehicle_Applications

[2] Jens Behley et al., "SemanticKITTI: A Dataset for Semantic Scene Understanding of LiDAR Sequences," arXiv:1904.01416, 2019. [Online]. Available: https://arxiv.org/abs/1904.01416

[3] Jakob Geyer et al., "A2D2: Audi Autonomous Driving Dataset," arXiv:2004.06320, 2020. [Online]. Available: https://arxiv.org/abs/2004.06320

[4] Pei Sun et al., "Scalability in Perception for Autonomous Driving: Waymo Open Dataset," CVF, 2020. [Online]. Available: https://openaccess.thecvf.com/content_CVPR_2020/papers/Sun_Scalability_in_Perception_for_Autonomous_Driving_Waymo_Open_Dataset_CVPR_2020_paper.pdf

[5] Holger Caesar et al., "nuScenes: A multimodal dataset for autonomous driving," arXiv:1903.11027, 2020. [Online]. Available: https://arxiv.org/abs/1903.11027

[6] Bin Yang et al., "PIXOR: Real-time 3D Object Detection from Point Clouds," arXiv:1902.06326, 2019. [Online]. Available: https://arxiv.org/abs/1902.06326

[7] Hang Su et al., "Multi-view Convolutional Neural Networks for 3D Shape Recognition," arXiv:1505.00880, 2015. [Online]. Available: https://arxiv.org/abs/1505.00880

[8] Charles R. Qi et al., "PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation," arXiv:1612.00593, 2017. [Online]. Available: https://arxiv.org/abs/1612.00593

[9] Marius Cordts et al., "The Cityscapes Dataset for Semantic Urban Scene Understanding," arXiv:1604.01685, 2016. [Online]. Available: https://arxiv.org/abs/1604.01685

[10] Andreas Geiger et al., "Are we ready for autonomous driving? The KITTI vision benchmark suite". [Online]. Available: https://www.cs.toronto.edu/~urtasun/publications/geiger_et_al_cvpr12.pdf