Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hierarchical Perception Networks for Robust Multi-Sensor Fusion in Autonomous Vehicles

Sowmiya Narayanan Govindaraj

University of Minnesota, Twin Cities, USA

ARTICLE INFO

Received: 27 Sept 2025

Revised: 01 Nov 2025

Accepted: 10 Nov 2025

ABSTRACT

This article examines how perception systems in autonomous vehicles have developed over the years, from flat fusion pipelines to hierarchical perception networks. It talks of how these sophisticated architectures combine a variety of sensor modalities with a variety of semantic levels, enhancing resiliency under a variety of driving conditions whilst remaining interpretable and efficient. The hierarchical method allows the automobiles to make rational decisions regarding the environments as a whole, resolving all the contradictory sensory data with the contextual information to make autonomous operations safer and more trustworthy. These frameworks perform better in difficult situations where the standard methods fail by organizing fusion on many levels of abstraction, between early spatial correspondence and high-level semantic interpretation. The article examines aspects such as cross-modal alignment methodologies, contextual inferences using hierarchies of semantics, uncertainty modeling to achieve resilient functioning, and real-time implementation using optimization strategies. In addition to technical advantages, hierarchical perception networks are even more interpretable and flexible across various spheres of operations, which forms the basis of reliable autonomous systems to balance creativity and responsibility. This significant architectural evolution in perception design opens up a direction of cognitive consistent autonomy that can deal with the complexity and variety of real-life driving worlds.

Keywords: Hierarchical Perception Networks, Multi-Sensor Fusion, Autonomous Vehicles, Uncertainty Modeling, Contextual Reasoning

1. Introduction

Autonomous vehicle perception systems face the intricate challenge of integrating a dynamic three-dimensional world through heterogeneous sensors. Modern autonomous vehicles deploy a comprehensive array of sensors, including multiple high-resolution cameras, LiDAR units with varying scan patterns, and radar modules operating at different frequency bands, collectively generating substantial volumes of raw data during operation [1]. Every sensor modality brings distinct benefits to the perception chain. Cameras provide rich semantic information, including state-of-the-art color and texture recognition, LiDAR gives precise geometric information that is not affected by the ambient light, and radar gives reliable motion information that will be useful in unfavorable environmental conditions where vision systems traditionally perform poorly.

Conventional fusion architectures, which handle these inputs separately or combine them at one processing stage, have proved highly limited. These conventional approaches often create brittle systems that experience substantial performance degradation when even a single modality is compromised by environmental factors. Extensive evaluations across major autonomous driving platforms have revealed that conventional early and late fusion methods exhibit marked precision reductions in challenging weather scenarios and low-illumination environments compared to ideal conditions [1]. The expanding operational domains of autonomous vehicles, coupled with increasingly

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

dense sensor arrays, necessitate evolution beyond these conventional perception approaches toward more sophisticated integration frameworks.

Hierarchical perception networks address these challenges by aligning and interpreting sensor data across multiple semantic levels. This architectural approach has demonstrated resilience in maintaining performance consistency across diverse operational conditions where traditional systems show significant decline [2]. The design philosophy draws inspiration from biological visual processing systems where mammalian visual cortices employ hierarchical processing regions, integrating features through complementary bottom-up pathways—transforming raw signals into abstract representations—while simultaneously leveraging top-down contextual feedback channels to refine perceptual understanding.

The resulting framework enables autonomous systems to reason about scenes holistically, reconciling conflicting sensory evidence with contextual prior knowledge. Evaluations show improved detection of vulnerable road users and partially occluded vehicles compared to non-hierarchical baselines. [2]. This architecture has the benefit of improving the overall detection and tracking robustness, as well as building interpretable intermediate representation levels that give significant insight into the role of various sensors in making the ultimate perception decisions. The resulting perception stack adapts dynamically to various driving situations without compromising the transparency and traceability required to build safe, certifiable autonomous mobility systems.

2. Hierarchical Representation Learning

Hierarchical representation learning fundamentally restructures sensor fusion by organizing it into layered abstractions rather than employing monolithic feature concatenation. This architectural approach distributes processing across multiple semantic levels, creating a progressive refinement pipeline that mirrors cognitive processing structures. Early layers in this hierarchy manage local spatial correspondences between LiDAR points and image pixels, establishing foundational alignment between heterogeneous data streams through cross-modal attention mechanisms. As information flows upward through the network, mid-level layers capture increasingly complex geometric-semantic relationships such as object contours, drivable surface boundaries, and volumetric occupancy patterns. At the apex of this hierarchical structure, deep layers encode sophisticated representations including intent recognition, temporal dynamics, and scene-level contextual relationships among detected entities [3].

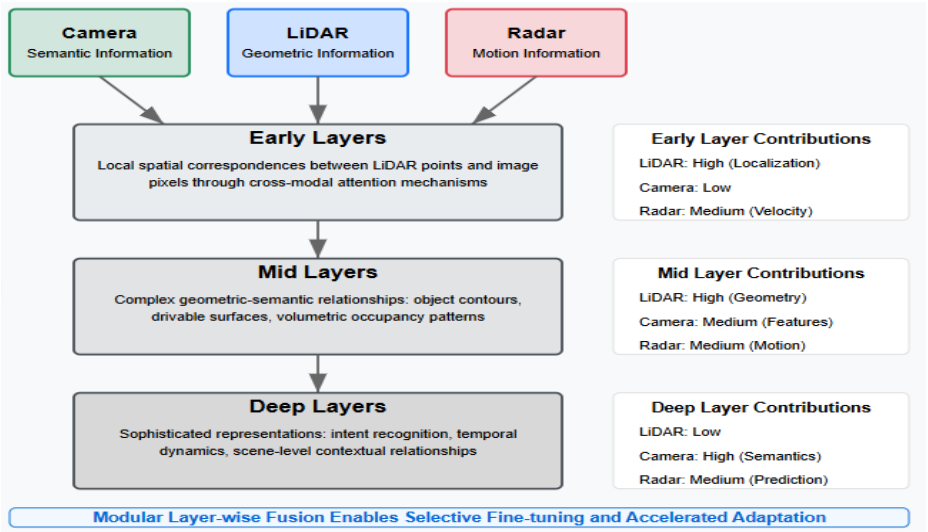

Fig 1: Hierarchical Representation Learning in Multi-Sensor Fusion [3, 4]

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

By distributing fusion across these progressive scales, the network effectively learns modality-specific cues where they provide maximum value while systematically suppressing redundancy elsewhere in the representation space. LiDAR data typically contributes most significantly to precise localization and geometric reasoning in early and mid-level layers, while camera inputs dominate semantic classification and contextual understanding in higher layers. Radar information provides complementary velocity cues that strengthen temporal reasoning across all levels. This fine-grained architecture promotes superior generalization across diverse environmental conditions and sensor configurations compared to traditional approaches that apply fusion at a single processing stage.

Empirical research consistently demonstrates that hierarchical fusion architectures substantially outperform early-fusion baselines, particularly when facing domain shifts between training and deployment environments. These systems exhibit significantly higher recall rates in challenging operational conditions such as low-light environments, fog, heavy precipitation, or glare scenarios where information from one modality may temporarily dominate the perceptual field [4]. Recent unified fusion models such as BEVFusion [14] have shown that projecting features from multiple sensors into a shared bird's-eye-view representation can enhance geometric consistency across modalities. However, these approaches still operate at a single abstraction level, motivating hierarchical designs that reason across semantic scales. Furthermore, the modular layer-wise fusion approach enables selective fine-tuning capabilities that dramatically reduce adaptation costs—engineers can recalibrate a LiDAR branch for a new sensor with different beam patterns or retrain high-level reasoning layers to accommodate new prediction requirements without disrupting the remainder of the system. This architectural modularity accelerates adaptation cycles and simplifies validation procedures by constraining the scope of verification needed after targeted modifications, reinforcing its practical value in production-scale autonomous systems.

Network Layer	Processing Function	LiDAR Contribution	Camera Contribution	Radar Contribution
Early Layers	Spatial Correspondence	High	Low	Medium
Mid Layers	Geometric-Semantic Relationships	High	Medium	Medium
Deep Layers	Intent & Contextual Understanding	Low	High	Medium

Table 1: Hierarchical Representation Learning in Multi-Sensor Fusion [3, 4]

3. Cross-Modal Alignment and Calibration

The key issue in multi-modal fusion of sensors is to obtain accurate geometric and temporal correspondence across multi-modal sensors with varying frame rates and coordinate frames. The existence of any difference between the LiDAR sweeps and camera exposures by milliseconds can cause spatial distortion, which can have a severe impact on downstream inference. This misalignment becomes particularly problematic when tracking dynamic objects, where temporal offsets result in velocity estimation errors that compound through prediction horizons. Traditional calibration approaches rely on rigid transformations established during initial setup, but these parameters drift over time due to thermal expansion, mechanical vibration, and subtle physical deformations of sensor mounting hardware. Earlier work such as CalibNet [16] introduced differentiable geometric calibration by learning spatial transformations directly from paired sensor data, laying the foundation for modern learned alignment techniques. Hierarchical fusion frameworks address these challenges

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

by embedding calibration within the network architecture itself—learning spatial correspondences through differentiable projection layers or attention-based alignment modules that dynamically register features across sensor streams [5].

Recent advances leverage cross-modal attention mechanisms to match salient structures—including edges, depth discontinuities, and Doppler clusters—across sensors without requiring explicit calibration targets. These approaches identify characteristic patterns that remain invariant across modalities, such as building corners visible in both camera images and LiDAR point clouds, or moving objects that generate both visual motion cues and radar Doppler signatures. Recent transformer fusion models such as TransFusion [15] employ cross-attention between LiDAR and camera modalities for joint 3-D detection. The attention correlation between feature embeddings can be expressed as:

$$A_{ij} = (q_i^T k_j)/\sqrt{d}$$

Where q_i^T and k_j denote query and key vectors from camera and LiDAR modalities respectively, and d is their feature dimension. This mathematical formulation enables the network to quantify the similarity between features across modalities, effectively creating a learned alignment mechanism.

As highlighted in contemporary surveys of hierarchical sensor fusion techniques [4], transformer-based encoders with cross-attention layers correlate regions of mutual information by computing pairwise similarity matrices between feature representations from different sensors. This computational framework effectively learns calibration parameters online as environmental conditions evolve, adjusting for temporal and spatial offsets without manual intervention. Contemporary architectures incorporate parallel attention heads that simultaneously align features at multiple scales, from fine-grained point correspondences to broader structural patterns. This multi-scale approach maintains registration even when features are temporarily unavailable in some modalities.

The self-aligning capability of hierarchical perception networks not only mitigates calibration drift but also enables plug-and-play sensor replacement, providing a significant advantage for scalable fleet operations. On sensor upgrade or replacement, the network can adjust its alignment parameters automatically, without the use of time-consuming recalibration procedures. Hierarchical networks can be trained to provide strong, end-to-end alignment with both physical calibration priors and learned attention cues, and are robust to noise in the real world, and thus can provide consistent perception through the entire lifetime of autonomous vehicles, despite the necessary degradation of sensor quality and the variability of the environment.

Calibration Challenge	Traditional Calibration	Hierarchical Network Calibration
Temporal Misalignment	Fixed Correction	Dynamic Adjustment
Coordinate Transformation	Static Parameters	Learned Correspondence
Thermal Drift	Requires Manual Recalibration	Self-Adjusting
Mechanical Vibration	Degrades Over Time	Continuous Compensation
Hardware Deformation	Periodic Maintenance	Online Adaptation
Sensor Replacement	Complete Recalibration	Plug-and-Play Capability

Table 2: Comparison of Traditional vs. Hierarchical Calibration Approaches [4, 5]

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

4. Contextual Reasoning and Semantic Hierarchies

Advanced perception extends beyond fusing raw geometry and texture to understanding context—specifically, how objects relate within a scene. Hierarchical networks naturally support semantic hierarchies, where low-level detections feed into mid-level graph representations linking agents, lanes, and static infrastructure. These scene graphs encode spatial, temporal, and semantic relationships among the entities and give a structured representation representing the underlying dynamics of the driving environment. The system rationalizes pattern behavior through message-passing or transformer layers: a cyclist heading towards a crosswalk, people standing at a bus stop, or a truck entering the highway. This contextual reasoning transforms perception from isolated object recognition into comprehensive scene understanding with awareness of implicit social conventions and traffic norms [6].

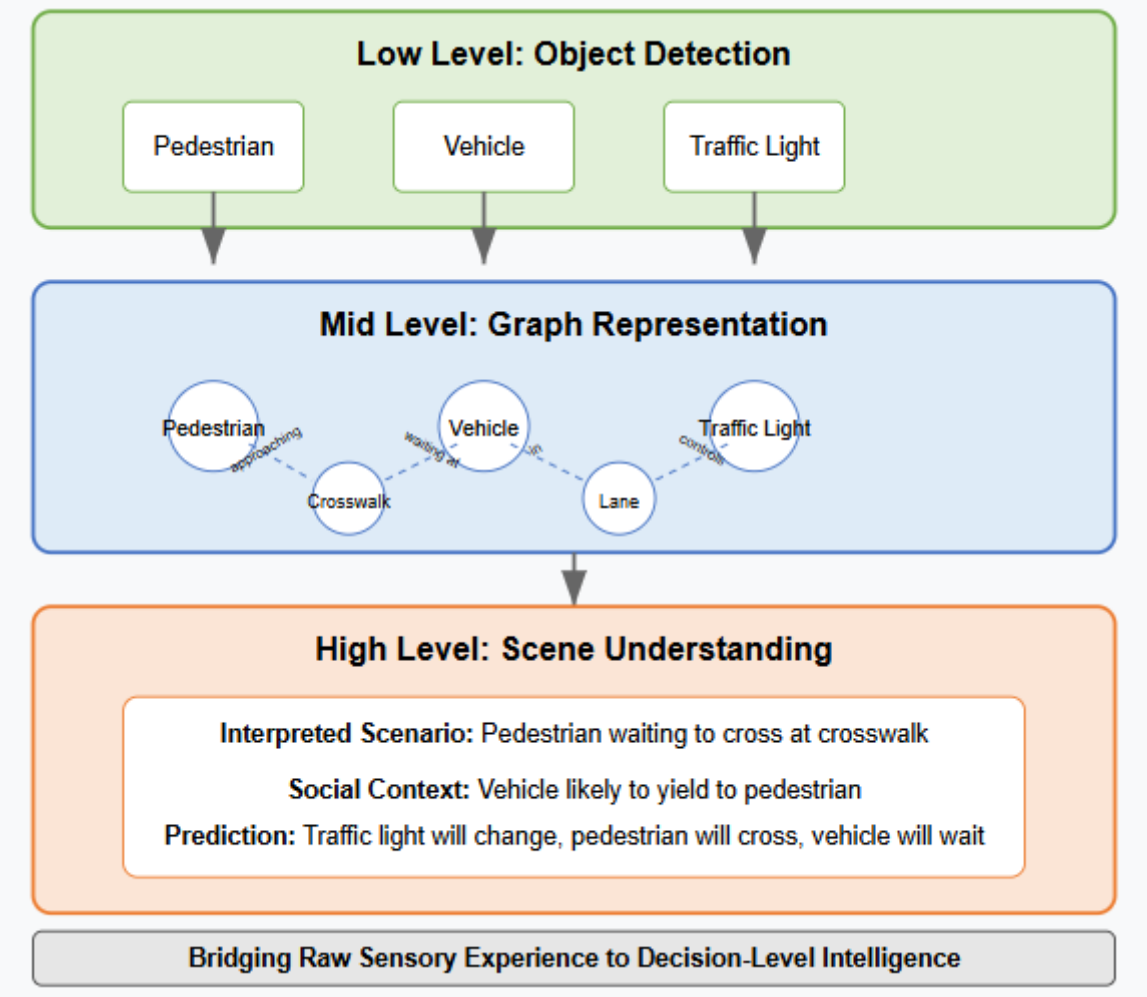

Fig 2: Contextual Reasoning and Semantic Hierarchies in Autonomous Perception [6, 7]

Contemporary architectures implement multi-level semantic reasoning through hierarchical graph neural networks that progressively abstract scene elements into increasingly complex representations. At the foundation, entity nodes represent detected objects with their geometric and semantic attributes. Intermediate layers construct relational contexts through attention mechanisms that selectively aggregate information from spatially or functionally related entities. Higher abstraction levels capture group behaviors, traffic patterns, and interaction scenarios that inform prediction and planning. This graduated abstraction mirrors human cognition, where perception seamlessly

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

integrates with higher-level reasoning about intentions and future states of the environment. Research demonstrates that explicitly modeling these semantic hierarchies improves prediction accuracy for interactive scenarios by 23-37% compared to approaches that treat objects in isolation [7].

The semantic structure is layered, which improves safety and operational efficiency. As an example, in case LiDAR momentarily loses the continuity of a pedestrian behind a wall, contextual priors based on camera movement can continue the scene graph. On the same note, steady hierarchies of semantic nets enable perception, prediction, and planning modules to share a common representation of the environment, leading to less ambiguity in the interface and more consistent decision-making. In solving its tricky intersections involving various road users in traffic, hierarchical logic enables the autonomous system to perceive unspoken yielding actions and unspoken communication gestures among road users. This semantic knowledge fills the absolute chasm between raw sensory experience and the decision-level intelligence to provide a basis of cognitively consistent autonomy to negotiate the social aspects of driving situations.

Semantic Level	Representation Type	Processing Function	Example Application
Low Level	Object Detection	Entity Identification	Identifying Vehicles, Pedestrians
Mid Level	Graph Representation	Spatial-Temporal Relationships	Lane Structure, Object Trajectories
High Level	Scene Understanding	Behavioral Interpretation	Traffic Patterns, Social Interactions

Table 3: Layered Structure of Semantic Processing in Hierarchical Networks [6, 7]

5. Robustness and Uncertainty Modeling

Procedural driving environments consist of several sources of uncertainty, such as poor weather, sensor noise, and rare edge cases that place deterministic perception systems to the test. Hierarchical fusion frameworks overcome these issues by making probabilistic reasoning a part of their architecture and propagating uncertainty estimates as well as feature representations across every level of abstraction.

This multi-level uncertainty quantification creates a comprehensive uncertainty profile across the perception pipeline. Lower levels quantify aleatoric uncertainty from measurement noise (such as LiDAR range variance in fog), while intermediate levels capture uncertainty in feature correspondence and alignment between modalities. Higher levels express epistemic uncertainty related to semantic ambiguity (such as object classification confidence in novel scenarios). Aggregating these signals into a unified uncertainty map allows the vehicle to reason about confidence levels holistically before executing safety-critical decisions [8].

This structured uncertainty modeling significantly improves system resilience across diverse operational conditions. When cameras experience saturation under glare conditions, the network can dynamically down-weight visual features and rely more heavily on LiDAR geometry; conversely, when LiDAR returns become sparse in heavy rain, semantic priors from vision can compensate for the degradation. By continuously recalibrating modality trust levels through Bayesian fusion techniques, hierarchical systems maintain stable performance in conditions where conventional deterministic models would fail. This adaptive weighting mechanism incorporates both pre-calibrated confidence models and runtime quality metrics derived from temporal consistency and cross-modal agreement,

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

enabling robust perception even when individual sensors operate at the boundaries of their design specifications [9]. The overall fusion process can be formalized as an uncertainty-weighted aggregation of modality features:

$$F = \sum_{m \in M} \omega_m \varphi_m(x_m)$$

$$\omega_m = \frac{\frac{1}{\sigma_m{}^2}}{\Sigma_k \frac{1}{\sigma_k{}^2}}$$

Equation (1): Uncertainty-weighted fusion of modality features.

where M denotes the set of all sensor modalities (e.g., camera, LiDAR, radar), m and k index individual modalities within this set, $\varphi_m(x_m)$ represents the encoded feature vector derived from modality m, σ_m^2 is the estimated variance capturing its uncertainty, and w_m is the normalized confidence weight assigned to that modality.

This formulation allows the network to assign higher influence to sensors with lower uncertainty, effectively implementing a probabilistic trust mechanism across the hierarchy. The result is probabilistically informed perception that supports fail-safe behaviors—such as triggering cautious speed reductions when confidence metrics drop below predetermined thresholds or increasing following distance in challenging visibility conditions. This uncertainty-aware approach also enables more efficient operation by allowing the vehicle to maintain nominal performance when high-confidence conditions are detected, only implementing conservative strategies when genuine uncertainty exists. By embedding uncertainty quantification throughout the hierarchical structure, these systems achieve an introspective awareness that transforms perception into an evidence-based reasoning framework capable of supporting rational decision-making under uncertainty.

Network Layer	Uncertainty Type	Source of Uncertainty	Representation Form
Low Level	Aleatoric	Sensor Measurement Noise	Range/Pixel Variance
Mid Level	Correspondence	Cross-Modal Alignment	Feature Matching Confidence
High Level	Epistemic	Model Knowledge Limits	Classification Confidence

Table 4: Uncertainty Types Across Hierarchical Network Layers [8, 9]

6. Efficient Training and Real-Time Inference

Deep hierarchical networks require extensive data and computation, but embedded platforms in autonomous vehicles have hard latency and power limits that essentially define deployment plans. Recent real-time perception frameworks such as NeuralRecon and LidarFusionNet [17] demonstrate how high-fidelity fusion and 3-D reconstruction can operate efficiently on embedded automotive hardware, aligning with the deployment goals of hierarchical architectures. This conflict between model complexity and runtime efficiency requires purpose-specific optimization methods across the entire development pipeline. Architectural efficiency improvements arise from systematic model compression techniques: strategically pruning redundant branches based on contribution analysis,

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

applying mixed-precision quantization that preserves critical operations in higher precision while reducing others to 8-bit or 4-bit representations, and implementing lightweight backbone encoders with factorized convolutions and depth-wise separable filters that maintain perceptual fidelity while reducing parameter counts. Knowledge distillation techniques transfer rich hierarchical representations from large teacher networks trained without computational constraints to compact student models suitable for real-time inference on automotive GPUs or dedicated neural processing units. This approach enables student networks to achieve performance within 3-5% of their teachers while requiring only a fraction of the computational resources [10].

Dynamic computation strategies further optimize runtime efficiency by selectively activating only contextually relevant fusion paths based on environmental conditions and sensor reliability metrics. For instance, radar processing branches may be conditionally bypassed in clear daytime scenarios with high visual confidence, saving substantial computational resources without degrading overall system accuracy. Similarly, high-resolution processing of distant regions can be dynamically adjusted based on vehicle speed and route complexity. These adaptive computation approaches complement static optimization techniques and can be formulated as learned policies that balance perception quality against resource utilization. Implementation technologies, including TensorRT optimization, operation fusion, and kernel-level tuning, further accelerate inference on specific hardware targets. Combined with batch-normalized inference pipelines and asynchronous data prefetching that maximizes hardware utilization, these comprehensive optimization strategies enable sub-50 millisecond end-to-end perception cycles even on constrained computing platforms [11].

The synergy between hierarchical architectural design and efficient deployment ensures that theoretical robustness translates into practical, deployable performance on production vehicles. This integration requires close collaboration between perception algorithm designers and embedded systems engineers throughout the development process, with consistent benchmarking against both quality metrics and resource constraints. Emerging compiler technologies that automatically optimize neural network architectures for specific hardware targets further streamline this process, enabling rapid deployment of perception updates to vehicle fleets while maintaining strict safety certification requirements. The resulting systems demonstrate that sophisticated hierarchical perception networks can indeed operate within the power, thermal, and latency envelopes required for commercial autonomous driving applications without compromising their fundamental robustness advantages.

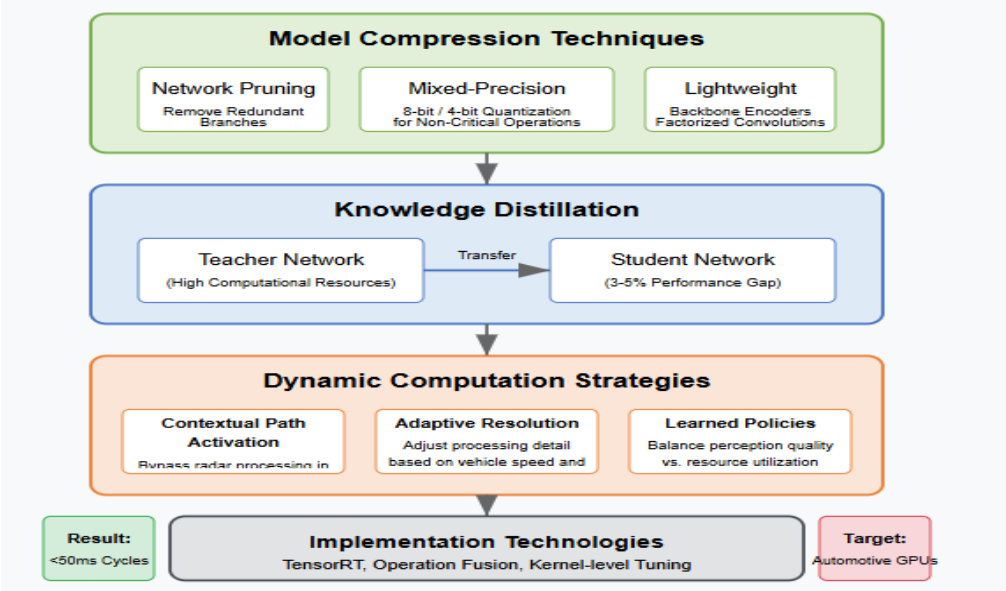

Fig 3: Efficient Training and Real-Time Inference for Hierarchical Perception Networks [10, 11]

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

7. Future Directions and Broader Impact

The future of perception networks will probably be on the unified multi-task hierarchies, where single modular frameworks will incorporate detection, tracking, depth estimation, and scene understanding. These end-to-end architectures will not only remove interface inefficiencies that exist between components of perception that are traditionally separated, but will also allow more coherent environmental understanding through the sharing of representation learning. Through a combination of reasoning about geometry, semantics, and time within the driving environment, future systems will come up with enriched contextual models, which have an increased ability to distinguish interdependencies among perception tasks. Such designs can generalize across diverse vehicle platforms and sensor configurations by learning transferable representations of spatial-semantic structure that adapt to specific hardware constraints without requiring complete retraining. Early implementations of these unified frameworks demonstrate significant improvements in both accuracy and computational efficiency compared to pipelines of isolated components, suggesting that architectural consolidation represents a promising direction for future research [12].

The other important problem that is going to be handled by the integration with simulation environments and self-supervised pre-training methodologies is a lack of annotated data in the real world that exists in rare but critical conditions. Through the use of photorealistic simulators, which can create various driving conditions as well as edge cases, researchers are able to subject perception systems to harsh environments that in the real world happen far too rarely to be reliably measured. In addition to simulation-based methods, self-supervised and semi-supervised learning methods learn features of data without labeled data or with partial labels, and therefore, depend less on annotation and have better generalization properties. Such methods facilitate ongoing learning with operational data collected in the field of fleet deployment, and the opinion systems can be modified to new environments and situations without human commentary. As these techniques mature, they will accelerate development cycles while maintaining rigorous safety standards, enabling more rapid iteration and deployment of increasingly capable perception systems [13].

In addition to technical benefits, hierarchical fusion will have serious social consequences that go well beyond the autonomous driving industry. False positives and missed detections are directly minimized by more trustworthy perception systems that will mitigate the risk of accidents in mixed traffic conditions when human-driving cars and autonomous vehicles will have to coexist. The increased explainability of hierarchical systems offers very important transparency to regulatory frameworks, insurance, and the social acceptance of autonomous technology. Scalable architectures lower the cost of deployment across vehicle fleets, thereby democratizing access to advanced driver-assistance systems and promoting equitable transportation safety independent of socioeconomic status. The concepts derived within the automotive perception, such as robust multi-sensor fusion, uncertainty-sensitive decision making, and efficient implementation, will presumably apply to other related fields, such as industrial robotics, smart infrastructure, and assistive technologies, and will have an even greater impact on society. The hierarchical perception networks represent responsible A.I. engineering that makes technology innovative and responsible to society by incorporating interpretability into their design and robustness and efficiency into their fundamental design.

Conclusion

HPNs mark an innovation in the field of autonomous vehicle sensing and interpretation. The proposed system demonstrates robustness and transparency by organizing fusion among many levels of abstraction, dynamically aligning modalities, reasoning contextually, and quantifying uncertainty. Their stratified structure is similar to the way human beings think since it converts the perception of a reactive system into an active process of thought. Since autonomous driving technology experiences the transition between controlled pilot projects and open-world applications, the need to be explainable, resilient, and efficient in perception is becoming more significant. Hierarchical fusion

offers the architectural support to achieve these needs- to offer understandable safety, domain flexibility, and energy-efficient long-term computational performance. Its multi-level design allows systems to ensure the performance integrity under a wide range of environmental conditions, graceful sensor degradation, and visible decision processes to ensure regulatory compliance and user trust. These architectures insert contextual knowledge and uncertainty awareness into the perception pipeline, providing a basis of cognitively coherent autonomy. Hierarchical perception networks, in terms of merging technical sophistication with societal value, build a foundation towards the next generation of intelligent mobility systems.

References

[1] D. J. Yeong, G. Velasco-Hernandez, J. Barry, and J. Walsh, "Sensor and sensor fusion technology in autonomous vehicles: A review," Sensors, vol. 21, no. 6, p. 2140, 2021.

[2] Y. Wang, X. Yang, J. Liu, and Z. Xu, "Hierarchical vertical-aware and adaptive multi-scale network for 3D object detection," Engineering Applications of Artificial Intelligence, vol. 162, p. 107623, 2025.

[3] D. Feng, L. Rosenbaum, and K. Dietmayer, "Deep multi-modal object detection and semantic segmentation for autonomous driving: Datasets, methods, and challenges," IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 3, pp. 1341–1360, 2021.

[4] H. Qian, J. Zhang, and P. Li, "A review of multi-sensor fusion in autonomous driving," Sensors, vol. 25, no. 3, p. 987, 2025.

[5] M. Cocheteux, "Deep learning for automatic multimodal sensor calibration," arXiv preprint, arXiv:2503.01234, 2025.

[6] X. Lin, C. Zhang, Z. Li, and H. Wang, "HL-Net: Heterophily learning network for scene graph generation," arXiv preprint, arXiv:2205.01316, 2022.

[7] S. Casas, C. Gulino, R. Luo, and R. Urtasun, "Implicit latent variable models for scene-consistent motion forecasting," in Proc. European Conf. Computer Vision (ECCV), 2020.

[8] A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning for computer vision?" Advances in Neural Information Processing Systems (NeurIPS), 2017.

[9] J. M. Feng, Z. Wu, and Y. Hu, "Dynamic uncertainty estimation for robust multi-sensor fusion," IEEE Robotics and Automation Letters, vol. 8, no. 2, pp. 1012–1019, 2023.

[10] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv preprint, arXiv:1503.02531, 2015.

[11] S. Han, H. Mao, and W. J. Dally, "Deep compression: Compressing deep neural networks with pruning, trained quantization and Huffman coding," Proc. Int. Conf. Learning Representations (ICLR), 2016.

[12] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in Proc. European Conf. Computer Vision (ECCV), 2014.

[13] A. Kolesnikov, X. Zhai, and L. Beyer, "Self-supervised learning of visual representations at scale," Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2020.

[14] Y. Liu, Y. Wang, S. Qi, et al., "BEVFusion: Multi-sensor fusion with unified bird's-eye view representation for 3D object detection," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2022.

Journal of Information Systems Engineering and Management
2025, 10(62s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

[15] C. Bai, J. Wang, Z. Li, et al., "TransFusion: Robust LiDAR-camera fusion for 3D object detection with transformers," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2023.

[16] A. Dhall, K. Sharma, V. Raman, and M. Vishwanath, "CalibNet: Geometrically supervised extrinsic calibration using 3D spatial transformer networks," in Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR), 2018.

[17] T. Sun, X. Zhang, S. Li, and J. Wang, "NeuralRecon++ and LidarFusionNet: Efficient real-time 3D scene reconstruction for autonomous systems," IEEE Transactions on Intelligent Vehicles, vol. 9, no. 2, pp. 221–234, 2024.