Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 17, No. 3, 2026

Perceptual Incompleteness in Autonomous Driving: From Information Deficiency to System-Level Propagation

Yushen Lin, Junyang Zhao *, Yaru Li, Yuxuan Li and Yutie Wang

Laboratory of Intelligent Control, Rocket Force University of Engineering, Xi'an, Shaanxi, 710025, China

*** Corresponding author:** Junyang Zhao (Email: zjy666sense@outlook.com)

Abstract: Autonomous driving perception systems face inherent observational incompleteness due to sensor field-of-view constraints and mutual occlusion among targets. The prevailing supervised learning paradigm, which presupposes fully annotated data, fails to provide effective behavioral constraints for models in information-deficient regions. From an information-theoretic perspective, this paper defines perceptual incompleteness as an intrinsic limit imposed jointly by observational geometry and physical mechanisms, and accordingly establishes a three-fold taxonomy comprising spatial, temporal, and causal deficiency. For each type of deficiency, the paper examines its origins, essential characteristics, and existing mitigation strategies. On this basis, the paper examines the systemic limitations of existing approaches from four perspectives: causal coupling, interface attenuation, semantic deviation, and evaluation blind spots. The analysis reveals that the three types of deficiency amplify each other through physical coupling; the perception-to-prediction and prediction-to-planning interfaces filter out uncertainty information through probability binarization and tail branch truncation, respectively; there is no direct mathematical correspondence between perceptual confidence and planning risk probabilities; and current evaluation frameworks systematically exclude information deficiency from the assessment scope. To address these limitations, this paper proposes four improvement strategies: constructing a joint modeling framework to address the causal coupling among the three types of deficiency, designing probabilistic module interfaces to prevent information attenuation, establishing standardization of cross-module probabilistic semantics, and developing a closed-loop risk-oriented evaluation system. The four improvements target training, interface transmission, semantic interpretation, and performance validation respectively, forming a complete solution from algorithm design to system integration to effect validation.

Keywords: Perceptual Incompleteness; Uncertainty Propagation; Modular Architecture; Autonomous Driving.

1. Introduction

Autonomous driving systems make decisions on the basis of what they can perceive. Perception is therefore not only an upstream recognition task: it sets the information boundary within which prediction and planning operate. For a driving system to act reliably, its environmental representation should preserve geometric relations, semantic meaning, and temporal continuity. Deep learning has substantially improved perception performance on annotated benchmarks. Yet difficult cases remain common in practice, particularly when an object is occluded, falls outside sensor coverage, or may appear only in the future. In such cases, the problem is not simply lower accuracy. The system often has limited evidence for determining what exists, what may happen next, and how much confidence should be assigned to its own output. The prevailing supervised-learning setup is not well matched to this situation. It is usually trained on labeled observations and optimized to produce a definite result for each input. This works reasonably well when the relevant scene content is visible and represented in the data. Real traffic scenes do not consistently meet that condition. Field-of-view limits create areas that cannot be observed, while vehicles, pedestrians, and roadside structures repeatedly block one another. A system may consequently need to act where the observation does not uniquely determine the scene state. This distinction matters for safety. When evidence is incomplete, a prediction may reflect model initialization, learned priors, or incidental correlations in the training set as much as it reflects the current environment. A high-confidence output is not necessarily a well-supported output. Supervised learning is designed to maximize predictive accuracy over annotated examples; an autonomous vehicle, however, must retain a defensible basis for action when information is missing. The gap between these objectives is most visible in information-deficient regions, where incorrect judgments can propagate to later stages of the driving stack. For this reason, perception under incomplete observation should be examined as a system problem rather than solely as a module-level accuracy problem.

In driving scenes, incomplete observation takes several qualitatively different forms. A vehicle may recognize that an area exists behind an occluder or beyond sensor coverage while remaining unable to determine what occupies it. This is referred to here as spatial deficiency. A different limitation concerns future traffic evolution: trajectories, interactions, and potential conflicts matter for the current driving decision, but the events themselves have not yet taken place. This constitutes temporal deficiency. There is also a less directly visible source of uncertainty. Intent, social conventions, and interaction strategies influence behavior, yet they are not observable in the same way as position or velocity. Inferring these hidden mechanisms from an observed action or trajectory gives rise to causal deficiency. These cases involve different reasoning tasks. For an occluded or out-of-view area, the issue is the absence of state information at a recognizable location. Future evolution cannot be recovered as if it were an already existing state; it must instead be represented through

alternative possibilities and their probabilities. Behavioral motivation presents another challenge, because observed outcomes alone may be compatible with several latent explanations. Their different information-theoretic properties and modeling demands make a unified treatment inadequate in many cases.

A number of studies have addressed parts of this broader problem. FastTracker[1] uses an occlusion-adaptive tracking mechanism and motion-model extrapolation to preserve target identity during short occlusions. MatchInformer[2] jointly encodes visible trajectories and occluded areas, drawing on clues in observable regions through a Transformer architecture to infer hidden states. Evidential Semantic Lane-Grid[3] distinguishes unknown regions semantically and updates occupancy probabilities through Bayesian inference. GMS [4] addresses future motion by introducing latent intent variables into conditional variational inference. Rather than committing to one path, it generates several trajectory candidates that may correspond to different behavioral choices. Vec-QMDP [5] places the problem in a hierarchical partially observable decision setting. Its focus is not only on forecasting possible state changes, but also on selecting actions while maintaining a belief over those changes. Drive-OccWorld [6] takes a different direction: it learns an implicit representation of the environment from vision foundation models, then uses latent-state transitions to examine possible scene changes and counterfactual outcomes. The three methods are not directly equivalent. They operate at different levels of the driving problem, from trajectory generation to belief-space decision-making and latent world modeling. They nevertheless share a practical purpose: providing a usable estimate when direct observation is incomplete.

Comparing their reported results is not straightforward. Each method is tested under its own task formulation, dataset split, input representation, and metric. A tracking or occupancy method may be assessed by state-estimation accuracy, whereas a prediction method is commonly judged by displacement error or coverage of future trajectories. Intent-related approaches introduce another difficulty because a single, observable ground-truth label may not exist for every behavior. As a result, the literature does not yet provide a common experimental basis for determining which strategy is more reliable across spatial, temporal, and causal forms of missing information. There is also a gap between component-level results and behavior in an integrated driving stack. Perception, prediction, and planning are often evaluated separately. During deployment, however, the output of one module becomes an assumption for the next. An uncertain occupancy estimate may be thresholded before it reaches prediction; several predicted motion modes may be reduced to one trajectory before planning; and uncertainty associated with an occluded object may be lost before its safety consequence is considered. The effect of these transformations is difficult to observe in isolated module experiments, although it is directly relevant to vehicle behavior in safety-critical scenes.

For this reason, the following discussion compares existing methods from three perspectives. The first concerns the status assigned to information that cannot be observed directly. Some methods assume that the missing state is short-lived and can be propagated from earlier observations. Others treat it as a persistent source of uncertainty that should remain in the scene representation. The second concerns uncertainty estimation. In particular, it is necessary to distinguish variability that is inherent in traffic behavior from uncertainty caused by insufficient observations or limited model knowledge. These correspond broadly to aleatoric and epistemic uncertainty, but the distinction is useful only when it affects the representation or the subsequent driving decision. The third perspective concerns the interface between modules. A probabilistic estimate has limited value if it is replaced by a point estimate before the next module can use it. These perspectives make it possible to discuss spatial, temporal, and causal deficiency without assuming that they are identical problems. They direct attention to what is missing, how the system records its uncertainty, and whether that information remains available after it is passed through the driving pipeline. Chapter 2 examines the physical origins and modeling treatments of the three deficiencies. Chapter 3 turns to system-level consequences, including their interaction, loss of uncertainty at module boundaries, inconsistent meanings assigned to probability values, and limitations of current evaluation practices. Chapter 4 then discusses joint modeling, probabilistic interfaces, shared probabilistic semantics, and closed-loop risk assessment as possible responses.

2. Conceptualization and Modeling of Three Observational Deficiencies

2.1. Basic Conceptualization

Unobservable information must be distinguished from concepts such as noise and false detections. Noise manifests as random perturbations to the true state, and its effects can be suppressed through filtering, robust estimation, or multi-frame accumulation; it therefore belongs to the domain of observation quality rather than information deficiency. False detections and missed detections reflect the system's failure to process already-available observational information, attributable to technical factors such as model capacity, training sufficiency, or algorithmic design. Unobservable information resides at a different level, originating from the physical constraints of the observation process rather than from limitations in data processing capability. Sensor fields of view inherently possess boundaries, and mutual occlusion among targets causes irreversible loss of spatial information. The above factors jointly determine that current observations are insufficient in the information-theoretic sense to uniquely determine the true state beyond occluded regions or outside the field of view. This deficiency is not statistical data sparsity but rather the inherent incompleteness of the observation mechanism itself.

The following three types of deficiency are distinguished based on differences in their physical origins and information-theoretic characteristics, rather than sensor types or data modalities. On this basis, Sections 2.2 through 2.4 sequentially present the conceptualization and modeling strategy analysis for spatial, temporal, and causal deficiency.

2.2. Spatial Deficiency

A sensor can reveal that a region exists without revealing what occupies it. This occurs when another object blocks the line of sight or when the region falls outside the sensor's field of view. Position, size, category, and motion state may all remain unknown. The safety implication depends heavily on whether the system treats this lack of evidence correctly. An area that has not been observed should not be treated as equivalent to an area observed to be free. Confusing the two can cause the vehicle to overlook a hidden obstacle; the

opposite error may lead to unnecessary braking or excessive conservatism.

Geometry gives spatial deficiency a relatively clear structure. The sensor viewpoint, the occluding object, and the scene layout constrain where the missing region lies and how large it may be. That structure changes as the ego vehicle, the occluder, and the hidden target move. A brief interruption in visibility may leave the target's state relatively well constrained, whereas a long interruption allows many possible states to develop. Any useful representation should capture both the location of the missing region and the history of the evidence that has been absent.

Recent surveys have documented the evolution of occupancy-grid methods from 2D bird's-eye-view grids to 3D occupancy estimation and 4D occupancy forecasting [7]. These developments highlight the importance of geometry-oriented representations for open-world driving scenarios involving long-tail obstacles. Vision-based 3D occupancy prediction has also become an active research direction, with recent reviews examining advances in feature enhancement, deployment efficiency, and label efficiency [8]. In parallel, multi-sensor fusion for BEV perception has been systematically reviewed for both single-vehicle and vehicle-to-everything scenarios [9].

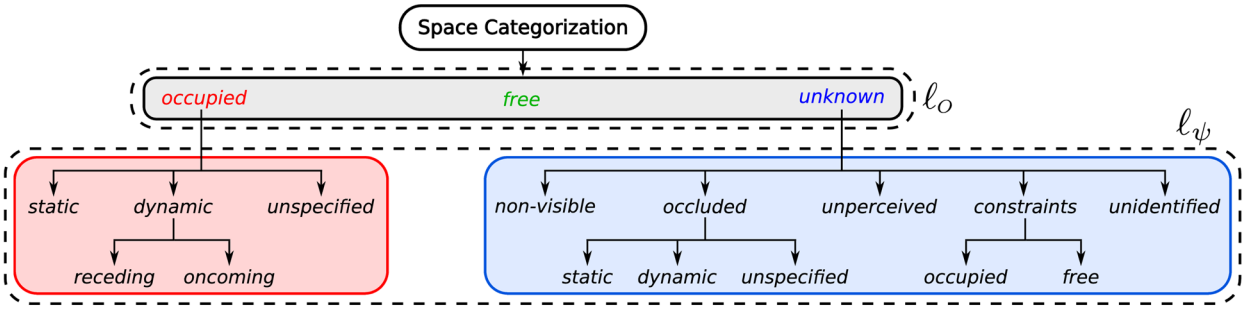

Fig 1. Space categorization based on DOG's estimation and potential causes of unknown space [3]

Several research directions make different use of this structure. FastTracker [1] continues an object's track through short occlusions by combining occlusion-aware processing with motion extrapolation. This strategy works when the target follows a sufficiently regular path, but uncertainty can grow quickly during a long disappearance. A single propagated track may eventually conceal several plausible states. MatchInformer [2] uses visible trajectories and neighboring scene information to infer the content of occluded regions through a Transformer-based architecture. Its predictions benefit from regularities in the training data, but the same reliance can produce confident errors when the scene distribution changes. Evidential Semantic Lane-Grid [3] takes a more explicit route. It divides the scene into grids, updates occupancy beliefs from current and historical observations, and distinguishes free, occupied, and unknown states. As illustrated in Fig. 1, this representation explicitly separates unknown space from free and occupied space while associating unknown regions with their potential causes.

The occupancy update can be expressed through Bayesian recursion:

$$P(\text{occ}\mid z_{1:t}) = \frac{P(z_t\mid \text{occ})\cdot P(\text{occ}\mid z_{1:t-1})}{P(z_t\mid z_{1:t-1})} \qquad (1)$$

where occ denotes the occupied state and z_t is the observation at time t. This recursion combines the likelihood of the current observation with historical beliefs to achieve sequential probability updates. These methods reflect different assumptions about risk. Track propagation often favors continuity and may become optimistic when an object behaves discontinuously after disappearing. Occupancy estimation preserves ambiguity but can mix sensor uncertainty, missing evidence, and model uncertainty within one probability value. In addition, an unknown label may cover situations with substantially different observation histories, such as a brief occlusion and a prolonged absence of evidence. Generative completion can produce detailed hypotheses, yet a plausible completion may be mistaken for an observed fact if the model does not provide provenance or confidence information. Thresholding creates an additional problem: a continuous occupancy belief may become a binary free/occupied label before prediction or planning receives it. Once that conversion occurs, later modules cannot reconstruct whether the original probability was 0.51, 0.85, or associated with a prolonged occlusion. A more informative interface should retain the estimated occupancy, the occlusion geometry, the duration of missing evidence, the visibility history, and an indication of whether the state came from direct observation, temporal propagation, or learned completion.

2.3. Temporal Deficiency

The future presents a different kind of information limit. A vehicle must often choose whether to accelerate, yield, stop, or change lanes before nearby agents reveal their next actions. The same current position and velocity can precede several different developments. A pedestrian may continue along the sidewalk, begin crossing, or stop in response to traffic. A vehicle may maintain its lane, merge, or yield. None of these outcomes exists as an observed state at the current time.

This distinction separates temporal deficiency from spatial deficiency. An occluded object has a physical state that exists even though the sensor cannot access it. A future maneuver has not yet occurred, so the system cannot recover it in the same way that it might infer an occluded position. Temporal prediction must instead represent a set of possible evolutions and estimate how each one relates to the available history and the candidate ego action. The Partially Observable Markov Decision Process (POMDP) provides a formal description of this setting. The agent maintains a belief over states, updates that belief with observations, and selects actions while accounting for uncertain transitions [10]. POMDP-based research has addressed localization, navigation, and interactive decision-making in autonomous driving and robotics [11].

Within the POMDP framework, the belief state can be written as:

$$b(s_t) = P(s_t \mid o_{1:t}, a_{1:t-1}) \tag{2}$$

where s_t is the current true state, $o_{1:t}$ is the observation sequence, and $a_{1:t-1}$ is the action sequence. The belief state represents a probability distribution over the current state conditioned on historical information. Its update depends on the observation model and the state-transition model. Because many future trajectories may be consistent with the same observation history, temporal prediction is inherently ill-posed. The objective is therefore not to recover an already existing but unobserved state, but to describe multiple possible evolutionary paths probabilistically.

The literature offers several ways to approximate this belief. MatchInformer [2] extrapolates future occupancy from historical context and temporal offsets. Its formulation is effective when the future resembles patterns in the training set, but it may struggle with unfamiliar interactions or distribution shifts. GMS [4] introduces latent intent variables into conditional variational inference and produces multiple trajectory hypotheses. This makes behavioral diversity visible, although the resulting modes may describe variation in the data rather than uncertainty about the present scene. Vec-QMDP [5] performs decision-making in belief space, as illustrated in Fig. 2, while UniAD [12] connects perception, prediction, and planning through a differentiable architecture. These approaches bring future prediction closer to decision-making, but they still depend on the content of the shared belief or latent representation. If an occluded agent, an unusual interaction, or a low-probability branch has already been removed, later optimization cannot restore it. Reviews of trajectory prediction have documented the continuing challenges of multimodality, interaction modeling, and uncertainty estimation [13].

A broad trajectory distribution does not automatically indicate that the model lacks knowledge. Its width may simply reflect ordinary behavioral variability. Conversely, a model may produce several modes and still remain overconfident if all of them come from familiar training patterns. This distinction matters at the planning stage. Averaging trajectories can hide meaningful differences between branches, while ranking or truncating hypotheses may discard an unlikely but dangerous outcome. A branch with a small probability can still dominate the decision if it creates a severe collision consequence under a particular maneuver. Temporal models should consequently record not only possible futures but also the evidence supporting them, the duration of any preceding observation gap, and the source of their uncertainty. Spatial and causal ambiguity further enlarge the prediction problem: an occluded agent has a less constrained current state, and the reason behind another agent's behavior may remain unclear.

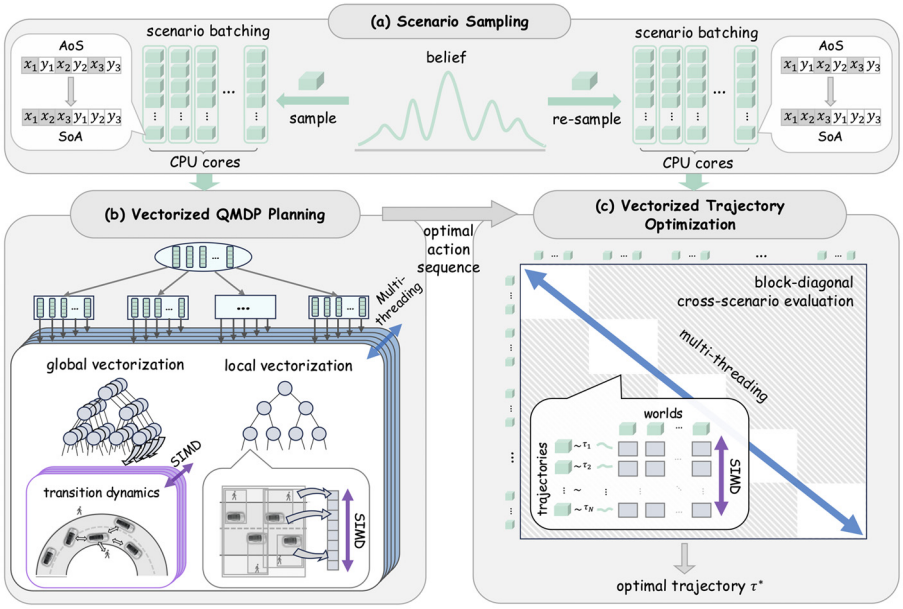

Fig 2. Overview of Vec-QMDP [5]

2.4. Causal Deficiency

Sensors reveal what road users do, but not necessarily why they do it. Position, velocity, posture, and scene geometry may indicate that a vehicle is slowing or that a pedestrian is approaching the curb. They do not directly disclose whether the driver intends to yield, merge, avoid an unseen obstacle, or change speed for some unrelated reason. Several behavioral mechanisms can produce similar observations, and one trajectory rarely identifies its underlying cause without additional assumptions. Causal deficiency concerns this gap between an observable outcome and the hidden mechanism that generated it. Let the observed behavioral trajectory be τ and the environmental state be s; then the intent z that the system attempts to infer and τ exhibit the following causal structure:

$$P(\tau \mid s) = \int P(\tau \mid s, z) \cdot P(z \mid s) dz \tag{3}$$

where z is the unobservable intent variable, $P(z \mid s)$ is the intent prior distribution, and $P(\tau \mid s, z)$ is the behavior generation model given intent. This expression indicates that the observed behavioral distribution is a mixture of behavioral patterns under multiple possible intents, and a single observed

trajectory τ cannot uniquely identify z. The core problem lies in inferring the hidden motivations that produced the observed behavior from the behavioral outcomes. The logical difference from the previous two types is that spatial deficiency concerns distinguishing between unknown and empty states, temporal deficiency concerns characterizing the upper risk bound of multimodal evolution, and causal deficiency concerns inferring latent intent from behavioral observations. Even with complete spatial information and accurate temporal prediction, if the system cannot understand why a behavior subject took a certain action, it still cannot make reliable judgments about its future behavior. The core issue concerns the interpretability and causality of latent variables; whether the implicit representations learned by the model correspond to genuine causal factors or merely fit statistical associations. The integral form above reveals a key difficulty: even if the system can fit the distribution $P(\tau|\ s)$, without additional causal assumptions or structural constraints, the decomposition of $P(z|\ s)$ and $P(\tau|\ s,z)$ is not unique; this is the non-identifiability problem in causal inference.

Different modeling strategies place the emphasis on interpretability, flexibility, or counterfactual reasoning. Fig. 3 presents the overall architecture of Drive-OccWorld [6], which performs latent-state transitions and counterfactual reasoning by combining intent-related information with spatial priors to evaluate multimodal planning trajectories. INTENT [14] assigns behavior to a finite set of categories, making the output relatively easy to interpret. Real intentions, however, may be continuous, ambiguous, and subject to revision during interaction. A driver can begin with a merging preference and switch to yielding when the available gap changes. GMS [4] introduces latent variables between scene observations and future motion, which allows the model to represent behavioral diversity without fixing a complete label vocabulary. The latent components may not correspond to genuine causes, however. They can absorb road geometry, dataset-specific correlations, or other statistical regularities. Their semantic meaning may also remain unclear to the planner. Drive-OccWorld [6] moves toward environment inference and counterfactual reasoning in a latent space. Generating alternative scene evolutions can support planning, but generation alone does not demonstrate causal understanding. A model may produce plausible futures without knowing which interaction mechanism separates one future from another. This weakness becomes more pronounced under distribution shift or unfamiliar social interactions. Reviews of pedestrian and vehicle behavior prediction have discussed the same difficulties in modeling intent, interaction, and behavioral uncertainty [16].

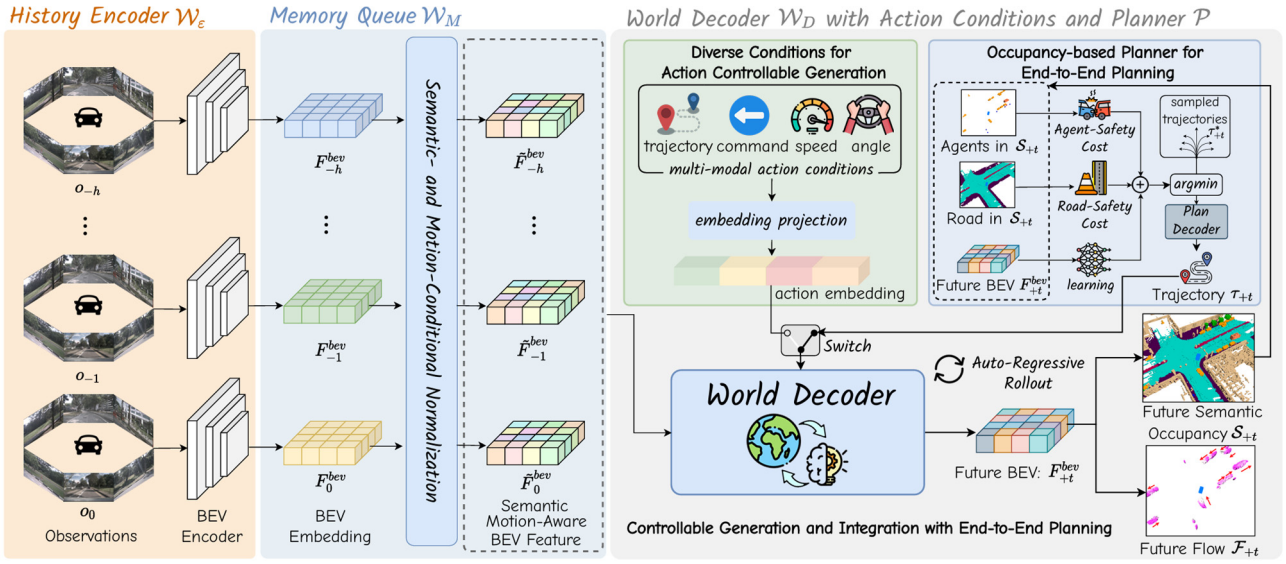

Fig 3. Overview of Drive-OccWorld[6]

Uncertainty in causal inference also requires careful interpretation. Aleatoric uncertainty reflects variation that remains even when the scene has been adequately observed. Epistemic uncertainty reflects missing evidence, limited model knowledge, or an unfamiliar situation [15]. The two cases may call for different actions. The planner can handle ordinary behavioral variation through risk-aware trajectory selection, whereas epistemic uncertainty caused by prolonged occlusion may justify additional observation or a larger safety margin. Causal inference also depends on the quality of spatial and temporal representations. If an unseen participant has disappeared from the scene model, the intent module may explain another agent's behavior without considering the interaction that caused it. If prediction removes a possible future branch, the corresponding behavioral explanation may never be evaluated. In ambiguous scenes, forcing the model to select one intent can be less reliable than retaining several explanations together with the evidence supporting each one. The purpose of causal modeling should not be to turn every uncertain behavior into a definitive label. It should help the system distinguish what the observed behavior supports from what remains conjectural.

3. Core Limitations of Existing Perception Approaches

Chapter 2 treated spatial, temporal, and causal deficiency as separate analytical categories. The driving stack does not experience them separately. A single incomplete observation can affect the current scene estimate, the set of predicted futures, and the interpretation of another agent's behavior at the same time. Consider a pedestrian who disappears behind a parked vehicle near a crossing. The missing position is a spatial problem. The possible reappearance locations create a

temporal prediction problem. The pedestrian's intention—waiting, crossing, yielding, or reacting to another unseen participant—remains causally ambiguous. Each uncertainty changes the others.

Table 1. Comparison of three types of observational deficiency

Dimension	Spatial Deficiency	Temporal Deficiency	Causal Deficiency
Physical Origin	Geometric occlusion and sensor field-of-view boundaries	The ontological fact that the future has not yet occurred	Inherent unobservability of behavior generation mechanisms
Core Problem	Distinguishing between "unknown" and "empty" states	Characterizing the upper risk bound of multimodal evolution	Inferring hidden motivations behind observed behaviors
Common Limitations	Multi-hypothesis merging; insufficient probability calibration; unreliable completion under distribution shift	Pattern fitting rather than causal inference; tail branch discarding; limited information from finite latent states	Discretized intent labels; latent prior mismatch; generative capacity ≠ causal understanding
Cross-Type Correlation	Spatial deficiency constitutes an amplification factor for temporal prediction uncertainty	Temporal deficiency constitutes a false premise for causal inference	Causal misjudgment leads to the breakdown of the prediction causal chain, with deviations reinforced in the closed loop

3.1. Causal Coupling of the Three Types of Deficiency

Conventional modular systems often pass a compact state vector from one stage to the next. Position, velocity, acceleration, and covariance may be sufficient when an object remains visible, but they rarely describe the reason for uncertainty. The receiving module may not know whether the covariance comes from sensor noise, a brief occlusion, a long period without evidence, or genuine variability in the agent's behavior. It may consequently process a weakly supported inference as if it were an ordinary noisy observation. PnPNet [17] connects perception and prediction within a differentiable pipeline, but a differentiable connection does not guarantee that the interface can express all decision-relevant uncertainty. If the perception stage sends only a narrow state representation, the prediction stage still cannot represent the full consequences of occlusion. Results from occlusion-aware prediction point in the same direction: incorporating occlusion reasoning and temporal consistency has been reported to reduce endpoint error for occluded agents by 25% and trajectory fluctuation by 10% to 20%. Streaming Motion Forecasting [18] retains several possible positions for an occluded agent rather than committing immediately to one path.

The problem extends beyond perception and prediction. A hidden scene element may also change the meaning of visible behavior. A partially observed pedestrian could be slowing because of an unseen vehicle, but an intent module that lacks this context may assign a precise explanation to an ambiguous maneuver. BeyondSight [19] identifies a related failure in end-to-end driving systems: an agent hypothesis may weaken or disappear when the agent is not continuously visible. Once that agent has been removed from the internal scene representation, the causal layer may construct an apparently complete explanation from incomplete participants.

Information can also flow in the reverse direction and reinforce an early mistake. Suppose the intent module interprets a yielding maneuver as a merge. The prediction stage may then favor trajectories that support acceleration and suppress the yielding alternative. Planning receives an optimistic future and has no clear signal that the original intent estimate should be reconsidered. A low-confidence spatial estimate, an incomplete temporal distribution, and an incorrect causal interpretation can thus form a self-reinforcing loop. The failure belongs to the interaction among modules rather than to one isolated prediction error.

This coupling does not make modularity itself undesirable. Modular systems remain valuable for engineering, testing, and deployment. The difficulty lies in interfaces that transmit selected state values while discarding the conditions and evidence behind them. Downstream components need to know whether an object was directly observed, propagated across an occlusion, inferred from contextual priors, or associated with several competing behavioral explanations. Without that information, the stack gradually replaces incomplete evidence with increasingly definite assumptions. Risk assessment should consequently examine how the three deficiencies interact across the complete driving pipeline, rather than evaluating each one only within its original module [20].

3.2. Information Attenuation at Module Interfaces

A perception model may internally maintain a rich representation of uncertainty. It can retain occupancy probabilities, several competing object states, covariance estimates, temporal evidence, and alternative interpretations of an incomplete scene. In many deployed pipelines, however, this information is simplified before being passed to the next module. One common operation is thresholding. A continuous occupancy probability is converted into an occupied or free label, while the unknown state may be treated as a low-confidence version of one of these categories. Another operation is hypothesis selection. Several possible object states or trajectories are replaced by the state with the highest estimated probability. These transformations reduce computational and interface complexity, but they also remove distinctions that may be important for safety. The perception-to-prediction interface is particularly vulnerable to this type of attenuation. If the perception module reports only a single position and velocity, the prediction module cannot determine whether the estimate was supported by continuous observation or inferred across a long occlusion. It may therefore produce a narrow trajectory distribution even when the upstream evidence is weak. A similar process takes place at the prediction-to-planning interface. Weighted averaging can turn several qualitatively different trajectories into one representative path, while truncation may eliminate low-probability branches before planning evaluates their consequences. UniAD [12] notes that independently deployed modules can introduce information loss, accumulated errors, and feature misalignment during cross-module processing. Joint perception-prediction research has reported related limitations in conventional modular systems, including limited sharing of computational resources, insufficient

integrated optimization, and weak propagation of uncertainty between components [21].

These issues are not merely implementation inconveniences. They determine which aspects of the scene remain available when the system reaches the decision stage. The loss is especially consequential for tail events. A trajectory with a small probability may nevertheless contain a high probability of collision under a particular ego action. Averaging such a trajectory with ordinary branches can reduce its visible effect, even though the planner should examine it precisely because of its consequence. Once the branch has been removed, the planning module cannot recover it from a deterministic input. This produces a gradual narrowing of the system's internal world model. Uncertainty may be represented in detail in the perception module, simplified at the perception-prediction boundary, averaged again by the prediction module, and finally presented to planning as a small collection of deterministic quantities. At each stage, the representation appears more stable and easier to process, but it also becomes less informative about what the system does not know. The final decision is then based on a world model that is not only incomplete because of the original observation conditions, but also incomplete because of the transmission process. The problem should not be interpreted as an argument against modularity in general. Modular architectures offer important advantages in engineering, testing, and deployment. The limitation lies in interfaces that assume deterministic state exchange even when the underlying task is probabilistic. A modular system can preserve uncertainty, but its interfaces must be designed to transmit the relevant alternatives and their associated confidence rather than only a selected output.

3.3. Cross-Module Deviation of Probabilistic Semantics

Even complete transmission of numerical uncertainty would not resolve the problem if different modules assign different meanings to the same value. Perception confidence, prediction likelihood, intent probability, and planning risk are related quantities, but they are not interchangeable. For perception, a confidence value may express the estimated probability that an obstacle occupies a particular location under the current observation, and its calibration depends on the model, data distribution, observation history, and target event definition. Prediction, in turn, may attach a weight to a trajectory mode, indicating how likely one future evolution is relative to other modeled alternatives. Planning, however, requires a different quantity: the probability of an undesirable consequence conditional on a candidate ego action and the relevant future state distribution—which depends on both future evolution and consequence models. Schematically, perceptual confidence concerns the probability that an obstacle exists at a given location given the observation history, whereas planning may need a conditional collision risk given a candidate action sequence and the same history; the second quantity depends on the chosen action, the evolution of surrounding agents, and their interaction with the ego vehicle. A high existence probability does not directly specify a collision probability, and a high trajectory likelihood does not necessarily imply high risk if the trajectory remains outside the ego vehicle's reachable set. Current systems may nevertheless use a confidence value as though it were a risk probability, or bypass probabilistic information altogether and make deterministic decisions. This creates ambiguity at the interface: when a prediction module assigns a value to a trajectory branch, the planning module may not know whether that value represents the probability of the complete trajectory, the confidence in a heading estimate, or the dispersion around a nominal path—each of which calls for a different planning operation.

Calibration research provides further reason for caution. Modern neural networks often produce scores that do not correspond closely to empirical posterior probabilities [22], so a numerical output may be internally consistent within one module while remaining unsuitable for direct interpretation by another. The issue becomes more difficult when the input distribution shifts under occlusion, rare interactions, or out-of-distribution behavior. UncAD [23] offers evidence that cross-module uncertainty can be useful when its meaning is preserved: by transmitting map uncertainty to planning, the system reportedly reduced collision rate by more than one-quarter and drivable-area conflict rate by more than two-fifths. These results support the value of uncertainty-aware transmission, but they also highlight the need for a defined semantic relationship between the transmitted quantity and the final planning objective. A system-wide probabilistic specification should therefore identify, for every uncertainty-bearing output: the event or variable to which the probability refers; the conditioning observations and time horizon; whether the value represents aleatoric or epistemic uncertainty; how multiple alternatives are normalized and retained; and which downstream operations are mathematically valid. Without these definitions, uncertainty may survive numerically while being lost conceptually—the system then appears probabilistic at the representation level but remains effectively deterministic at the decision level.

3.4. Systematic Blind Spots in Evaluation Frameworks

Evaluation protocols provide the final mechanism through which the research community decides whether a method has improved, yet existing protocols are still centered largely on visible targets and available ground truth. Metrics such as mean Average Precision, Intersection over Union, and occupancy accuracy are valuable for measuring recognition quality, but their definitions generally assume that the evaluated state can be labeled—an assumption not satisfied in many information-deficient situations. The nuScenes dataset [24] illustrates this emphasis on conventional recognition metrics, with mAP serving as a major criterion. Some newer occupancy benchmarks have begun to represent unknown, free, and occupied states more explicitly: SurroundOcc [25] provides dense 3D occupancy annotations for surround-view scenes, and OpenOccupancy [26] contributes to evaluating occupancy representations, yet these developments do not by themselves resolve the broader system-level problem. Prediction benchmarks offer another partial solution: the Waymo Open Dataset [27] and nuScenes prediction tasks evaluate future behavior using metrics such as minADE, FDE, and soft multimodal displacement, which recognize that future states are not directly observable. Nevertheless, average displacement remains an incomplete proxy for safety—a prediction can achieve low average error while omitting a rare branch that would have significant consequences for a particular maneuver. Three limitations are especially important. First, most metrics treat errors by numerical deviation rather than safety consequence: misclassifying an occluded region as empty versus labeling

an actually empty region as occupied may receive comparable treatment, yet the first may lead to an unsafe maneuver while the second only causes unnecessary braking. Second, offline metrics cannot fully capture uncertainty propagation through the complete software stack—a model may improve occupancy accuracy without transmitting the associated uncertainty to prediction, while a modest perception improvement may yield a substantial safety gain if it changes how the planner handles ambiguous regions (as the collision-rate reduction in UncAD [23] illustrates). Third, causal deficiency is rarely measured directly, since intent, interaction strategy, and behavioral motivation do not have a unique observable ground truth in many real scenes; a benchmark that requires one fixed intent label may conceal genuine ambiguity or penalize a model for maintaining multiple plausible explanations.

A more complete protocol should therefore operate at several levels to address these shortcomings. At the perception level, visible and information-deficient regions should be reported separately, together with calibration measures for the latter. At the prediction level, evaluation should include recall of rare or extreme events, preservation of high-consequence branches, and risk-weighted displacement measures. At the system level, closed-loop simulation or controlled testing should examine the relationship between reported confidence, selected action, near-collision behavior, and actual safety outcome, while scenario generators vary occlusion duration, sensor coverage, interaction patterns, and distribution shift so that performance is not inferred only from routine visible scenes. Such a multi-level protocol would provide a more faithful account of how uncertainty affects actual driving decisions, rather than rewarding models that merely fit labeled states without regard for what remains unseen.

Taking a broader view across Sections 3.1–3.4, these issues collectively reveal a fundamental gap between component-level competence and system-level reliability. Individual modules may perform adequately under their own evaluation assumptions, yet the complete stack can still fail when those assumptions are combined: spatial deficiency enlarges temporal uncertainty, interface compression removes the evidence needed to represent that uncertainty, semantic mismatch prevents planning from interpreting what remains, and conventional metrics fail to expose the resulting risk. The central architectural limitation is therefore not simply insufficient accuracy in one perception module—uncertainty is generated within modules but often discarded at interfaces, reinterpreted without a common semantic specification, and evaluated only indirectly. These four problems—causal coupling among deficiencies, information attenuation at interfaces, cross-module semantic deviation, and evaluation blind spots—are not independent but mutually reinforcing, and they constitute the core system-level limitation that motivates the improvement directions presented in Chapter 4, namely joint modeling, probabilistic interfaces, standardized uncertainty semantics, and closed-loop risk assessment.

4. Improvement Directions

The limitations discussed in Chapter 3 do not arise from a single weak module. They emerge when incomplete observations are processed by a pipeline whose components make separate assumptions about state, uncertainty, and risk. A practical response therefore has to address model construction, information exchange, semantic interpretation, and validation together. The four directions below are organized around those system functions rather than treated as interchangeable remedies.

4.1. Joint Modeling Framework

Spatial, temporal, and causal deficiency are commonly assigned to different tasks—detection and occupancy estimation resolve what is present, prediction estimates how the scene may evolve, and intent recognition explains or anticipates behavior. This separation is useful for implementation, but it becomes problematic when uncertainty in one task changes the reliability of another. For example, a pedestrian partly hidden by a parked vehicle should not be represented only by an uncertain position estimate; the duration and geometry of the occlusion also affect the credibility of a future trajectory forecast and of any inferred crossing intention. A joint framework should therefore preserve these dependencies during model training. One possible design uses a shared scene representation that contains occupancy uncertainty, alternative future evolutions, and latent behavioral explanations. Task-specific heads can still produce occupancy, trajectory, or intent outputs, but they should be conditioned on a common representation rather than on isolated module outputs.

The learning objective should likewise penalize inconsistent combinations. A trajectory branch that assumes an agent is visible, for instance, should receive reduced support when the spatial branch identifies prolonged occlusion; conversely, evidence of an interaction-sensitive maneuver may justify maintaining a broader occupancy or intent hypothesis. The purpose is not to force all uncertainty into one undifferentiated latent vector—spatial absence, future branching, and hidden intention remain distinct phenomena. The joint model should instead make their dependence explicit, so that one branch can modify the confidence assigned by another. In particular, occlusion duration and visibility history should be available to the prediction and intent components, since a short interruption has different implications from a long unobserved interval. Fig. 4 shows the UniUncer framework [28], which estimates uncertainty for static map elements and dynamic agents within one architecture, demonstrating that uncertainty associated with different scene entities can be handled in a common pipeline. UncAD [23] provides a related system-level result: online map uncertainty transmitted to prediction and planning reduced collision rate by up to 26% and drivable-area conflict rate by up to 42% with a 1.9% parameter increase. These studies do not solve every form of observational deficiency, but they indicate that uncertainty need not be confined to the module in which it first appears.

4.2. Probabilistic Module Interfaces

A joint representation is of limited value if its uncertainty is discarded at the next interface. In many driving stacks, perception outputs are reduced to a single state estimate before prediction begins; likewise, a multi-modal prediction is often condensed into one expected trajectory before it reaches planning. Such reductions are convenient for fixed-format interfaces, yet they cause the planner to act on a simplified scene in which unsupported alternatives have already been removed. The interface between modules should carry a distributional description of state rather than only a most-likely estimate. At a minimum, it should retain the estimated mean, covariance, hypothesis weights, and the

provenance of the uncertainty. Depending on the representation, this may take the form of occupancy distributions, weighted trajectory sets, samples, or parameters of a mixture model. The appropriate format is an implementation decision; the essential requirement is that downstream modules can distinguish a well-observed state from an inferred one and can account for plausible alternatives in their own update or planning process.

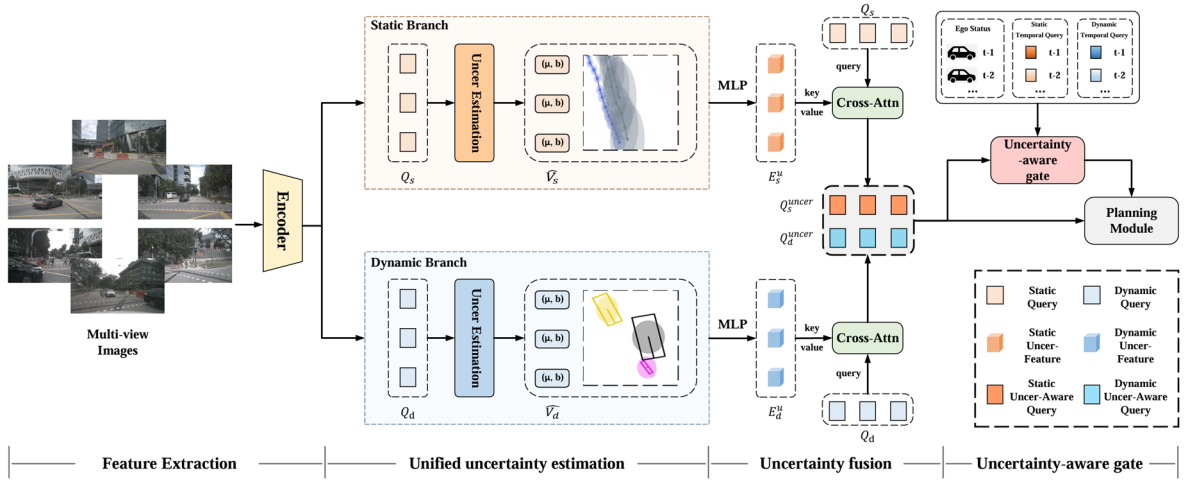

Fig 4. Overview of the UniUncer framework [28]

This changes the role of downstream modules. Prediction should treat perceptual output as a conditional prior rather than as ground truth; planning should evaluate actions against the relevant distribution of future outcomes instead of a trajectory produced by weighted averaging alone. Low-probability branches should not automatically disappear merely because they have little effect on a mean trajectory—their retention should depend on both likelihood and consequence, especially when a branch contains collision risk. Fig. 5 presents the uncertainty-propagation framework of Ivanovic et al. [29], which introduces perceptual state uncertainty (mean and covariance) into trajectory forecasting with a statistical-distance-based loss, identifying a common source of overconfidence: many predictors consume only the most likely upstream state and ignore its uncertainty. Related speculative-planning approaches show that behavioral and trajectory uncertainty can also be transferred across the prediction-planning boundary as distributions rather than point estimates. UncAD [23], which obtains substantial safety gains by propagating map uncertainty alone, further suggests that the design of interfaces deserves at least as much attention as the design of individual modules.

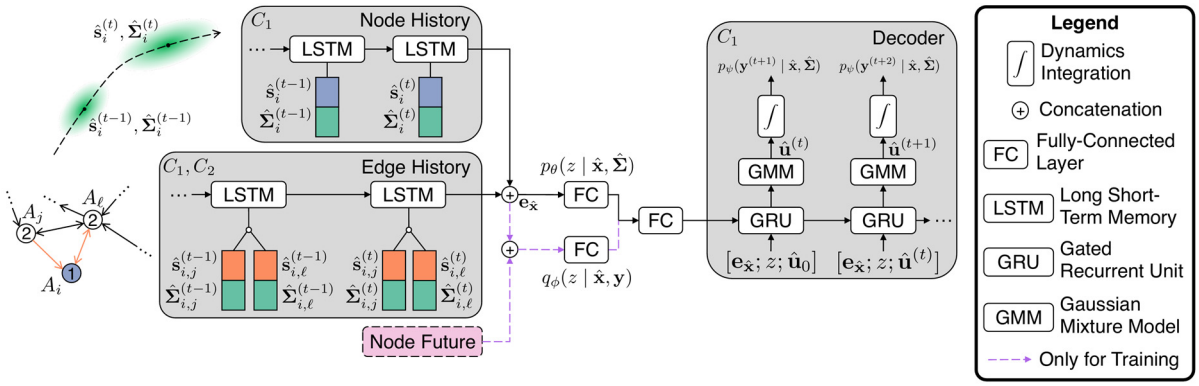

Fig 5. Overview of the uncertainty propagation framework [29]

4.3. Standardization of Probabilistic Semantics

Passing numerical uncertainty downstream does not by itself guarantee that it will be used correctly. A probability attached to an occupancy cell, a confidence score assigned to an intent class, and the probability of collision under a candidate action answer fundamentally different questions—the first concerns the state of a location, the second concerns a latent behavioral explanation, and the third depends on possible future states, the ego action, and their consequences. Treating these quantities as directly interchangeable creates a semantic error even when each value is individually calibrated. A system-level probabilistic specification is therefore needed to state what each module reports and how later modules may use it. Each output should identify its event definition, prediction horizon, conditioning information, uncertainty type, and calibration context. For example, a trajectory output should make clear whether its weight denotes the probability of a complete trajectory mode, uncertainty around a nominal path, or confidence in a model estimate. The planning module can then apply the appropriate transformation from upstream state uncertainty to action-conditioned risk.

This specification should also separate aleatoric and epistemic uncertainty where possible. Aleatoric uncertainty reflects variability that remains even with adequate knowledge of the scene, whereas epistemic uncertainty indicates insufficient evidence or limited model knowledge—these cases call for different responses. A planner may accept ordinary behavioral variation through risk-aware trajectory

selection, while epistemic uncertainty in a heavily occluded region may require information-gathering behavior or a more conservative safety margin. Calibration must be checked across the integrated pipeline, not only within individual modules; a perception model can be calibrated on its own output distribution while still providing values that a predictor or planner interprets incorrectly. Joint calibration tests should therefore examine whether stated probabilities retain their intended meaning after conversion between representations and after the system has selected an action. Fig. 6 illustrates how UncAD [23] estimates, propagates, and uses uncertainty across perception, prediction, and planning. Deglurkar et al. [30] similarly argue that uncertainty measures require interpretation in the context of system design and operation, rather than as universally comparable scalar scores. Reviews of uncertainty quantification for autonomous-driving perception [31] and deep learning more broadly [32] support this distinction, while deep ensembles [33] and evidential deep learning [34] offer complementary estimation mechanisms. The relevant design question is not which estimator is universally best, but whether the uncertainty it produces has a defined and usable meaning at the next decision stage.

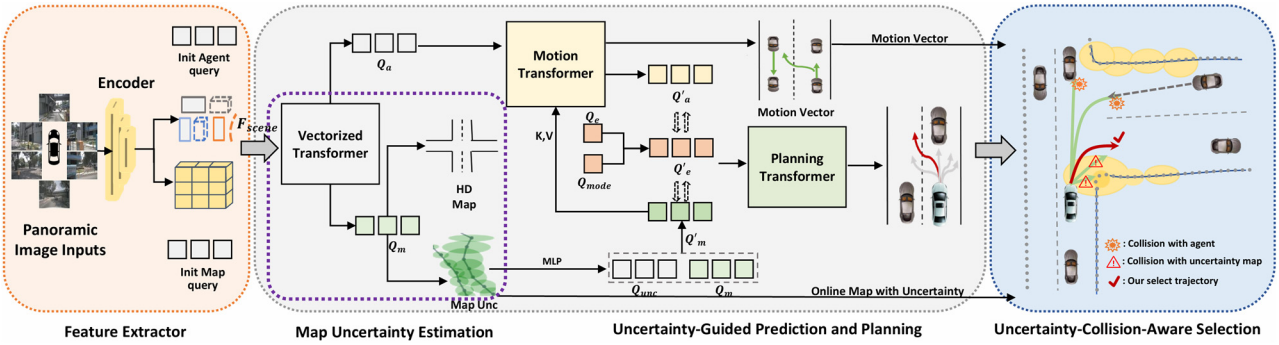

Fig 6. The overall architecture of UncAD consists of four core modules [23]

4.4. Closed-Loop Risk Assessment System

Existing benchmarks provide important measures of recognition and prediction accuracy, but they only partially expose failures caused by incomplete observation. Mean Average Precision, Intersection over Union, and occupancy accuracy are primarily defined around labeled states; as a result, an unobserved region may receive limited attention in evaluation even though an incorrect assumption about that region can determine the safety of a later maneuver. Offline gains in a perception metric also do not necessarily show whether uncertainty was preserved, interpreted properly, and used to prevent an unsafe action. A model that improves occupancy accuracy may still fail to transmit the associated uncertainty to prediction, while a modest perception improvement could yield a substantial safety gain if it changes how the planner handles ambiguous regions—a type of system benefit that is difficult to infer from isolated offline scores, as the collision-rate reduction reported for UncAD [23] illustrates. Evaluation should therefore connect module-level quality to closed-loop driving risk rather than stopping at component-level accuracy.

A more complete protocol should operate at several levels to address these shortcomings. At the perception level, visible and information-deficient regions should be reported separately, together with calibration measures for the latter. At the prediction level, evaluation should retain standard displacement metrics but add measures that reflect rare and consequential outcomes, such as extreme-event recall and risk-weighted displacement error—a model that accurately predicts routine behavior while systematically omitting a hazardous branch should not be regarded as adequate solely because its average error is low. At the system level, closed-loop simulation or controlled testing should vary occlusion patterns, visibility duration, interaction structure, and distribution shift while preserving reproducible scenario definitions, and should record not only collision frequency but also near-collision risk, inappropriate conservatism, intervention behavior, and the correspondence between reported uncertainty and the safety margin selected by the planner. Fig. 7 presents the SUPER-AD framework [35], which estimates aleatoric uncertainty in BEV space, produces an uncertainty-aware drivability map, and incorporates it into planning for closed-loop assessment. UncAD [23] likewise uses uncertainty in collision-aware planning after generating multi-modal trajectories. Together, these studies illustrate why evaluation cannot end with an offline score: the relevant outcome is whether the integrated system behaves more safely when its observations are incomplete.

Taking a broader view across Sections 4.1–4.4, the four directions above support one another, but their relationship is functional rather than a rigid sequence. Joint modeling makes dependencies among spatial, temporal, and causal uncertainty available during learning. Probabilistic interfaces prevent those dependencies from being collapsed into point estimates during deployment. Semantic standardization gives subsequent modules a basis for interpreting the transmitted information. Closed-loop evaluation tests whether the complete system converts that information into safer action. A weakness in any one part can undermine the others—for example, a joint model that preserves dependencies is useless if the interface discards them; an interface that transmits uncertainty is ineffective if downstream modules misinterpret its semantics; and even a semantically correct transmission cannot be validated if the evaluation protocol does not examine closed-loop safety consequences. For this reason, observational incompleteness should be managed as an end-to-end property of the driving stack rather than as a collection of isolated estimation errors, and the four improvements proposed here—targeting training, interface transmission, semantic interpretation, and performance validation respectively—constitute a complete solution from algorithm design to system integration to effect validation.

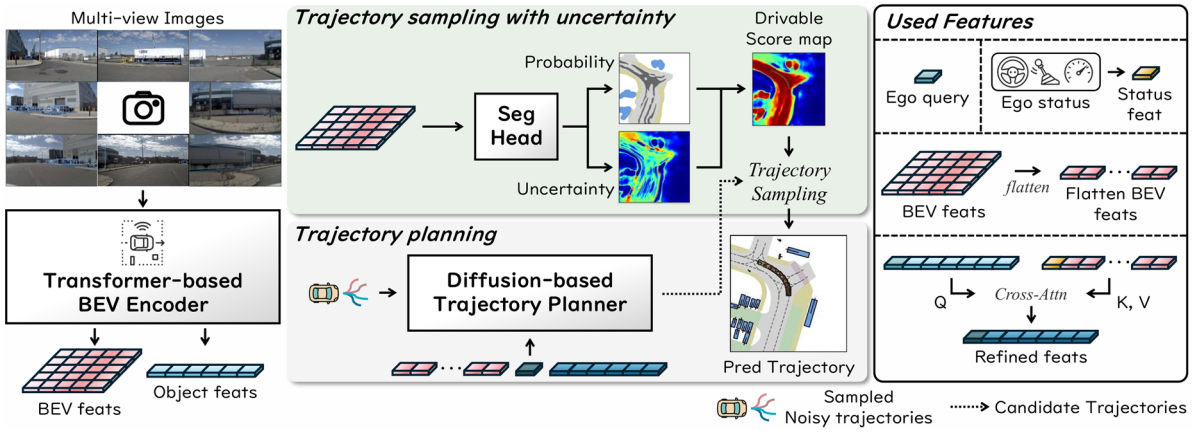

Fig 7. Overview of the SUPER-AD framework for uncertainty-aware planning [35]

5. Summary

This paper has distinguished three types of observational incompleteness in autonomous driving perception based on their physical origins. Spatial deficiency arises from geometric occlusion and field-of-view boundaries; temporal deficiency arises from the fact that the future has not yet become reality; and causal deficiency arises from the concealment of behavior generation mechanisms. For each type of deficiency, the paper has reviewed its core attributes and existing modeling strategies. Spatial deficiency is addressed through motion extrapolation, occupancy grids, and data completion; temporal deficiency through predictive extrapolation, feature collaboration, and closed-loop optimization; and causal deficiency through intent classification, latent variable modeling, and environment inference.

On this basis, the paper has analyzed the transmission mechanisms of uncertainty in the system when the above methods are integrated into a complete software stack. The analysis shows that causal coupling causes the three types of deficiency to amplify each other. The perception-to-prediction and prediction-to-planning interfaces filter out uncertainty information through probability binarization and tail branch truncation respectively, while the semantic gap between perceptual confidence and planning risk probabilities prevents even transmitted uncertainty from being correctly interpreted. Moreover, current evaluation frameworks systematically exclude information deficiency from the assessment scope, further obscuring the above failures.

The paper accordingly proposes four improvement strategies: a joint modeling framework, probabilistic module interfaces, standardization of probabilistic semantics, and a closed-loop risk assessment system. The four improvements target training, interface transmission, semantic interpretation, and performance validation respectively, thereby constituting a complete solution from algorithmic design to system integration to effect validation. It can be anticipated that breakthroughs in the above directions will drive the transformation of autonomous driving perception paradigms from precision-driven approaches toward frameworks capable of handling information incompleteness.

References

[1] Hashempoor, H., & Hwang, Y. D. (2025). FastTracker: Real-time and accurate visual tracking [Preprint]. arXiv. https:// arxiv. org/abs/2508.14370.

[2] Rothenhäusler, A., Mazzola, M., Look, A., et al. (2026). Don't double it: Efficient agent prediction in occlusions [Preprint]. arXiv. https://arxiv.org/abs/2601.21504.

[3] Jimenez-Bermejo, V., Trentin, V., Artunedo, A., et al. (2025). Evidential semantic lane-grid for unknown space analysis and high-level representation of dynamic occupancy grids. IEEE Transactions on Intelligent Vehicles, 10(7), 3993–4009. https:// doi. org/10.1109/TIV.2025.3482611.

[4] Chen, X., Gu, G., & Ning, X. (2026). GMS: Graph-guided multi-modal spatio-temporal fusion model for trajectory prediction in autonomous driving. Neurocomputing, 661, 132010. https://doi.org/10.1016/j.neucom.2026.132010.

[5] Ye, H., Xiao, Y., Lu, C., et al. (2026). Vec-QMDP: Vectorized POMDP planning on CPUs for real-time autonomous driving [Preprint]. arXiv. https://arxiv.org/abs/2602.08334.

[6] Yang, Y., Mei, J., Ma, Y., et al. (2025). Driving in the occupancy world: Vision-centric 4D occupancy forecasting and planning via world models for autonomous driving. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 39, No. 9, pp. 9327–9335). AAAI Press. https://doi.org/ 10. 1609/aaai.v39i9.32947.

[7] Shi, Y., Jiang, K., Li, J., et al. (2025). Grid-centric traffic scenario perception for autonomous driving: A comprehensive review. IEEE Transactions on Neural Networks and Learning Systems, 36(7), 11814–11834. https://doi.org/ 10.1109/ TNNLS. 2024.3429678.

[8] Zhang, Y., Zhang, J., Wang, Z., et al. (2026). Vision-based 3D occupancy prediction in autonomous driving: A review and outlook. Frontiers of Computer Science, 20(1), 13–34. https:// doi. org/10.1007/s11704-025-4112-8.

[9] Ping, P., Zhang, X., Tao, L., et al. (2026). A comprehensive survey on multi-sensor information processing and fusion for BEV perception in autonomous vehicles. Information Fusion, 126. https://doi.org/10.1016/j.inffus.2026.103487.

[10] Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. Artificial Intelligence, 101(1–2), 99–134. https://doi. org/10. 1016/ S0004-3702(98)00023-X.

[11] Kurniawati, H. (2022). Partially observable Markov decision processes and robotics. Annual Review of Control, Robotics, and Autonomous Systems, 5(1), 253–277. https://doi. org/ 10. 1146/ annurev-control-042920-091756.

[12] Hu, Y., Yang, J., Chen, L., et al. (2023). Planning-oriented autonomous driving. In 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 17853–17862). IEEE. https://doi.org/10. 1109/CVPR 52729. 2023. 01712.

[13] Xu, M., Liu, Z., Wang, B., et al. (2025). A survey of autonomous driving trajectory prediction: Methodologies, challenges, and future prospects. Machines, 13(9), 818. https:// doi. org/10.3390/machines13090818.

[14] Tang, Y., & Ma, W. (2026). INTENT: Trajectory prediction framework with intention-guided contrastive clustering. IEEE Open Journal of Intelligent Transportation Systems, 7, 337–352. https://doi.org/10.1109/OJITS.2026.3582164.

[15] Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In Advances in Neural Information Processing Systems (Vol. 30, pp. 5580–5590). Curran Associates.

[16] Galvão, L. G., & Huda, M. N. (2024). Pedestrian and vehicle behaviour prediction in autonomous vehicle system — A review. Expert Systems with Applications, 238. https://doi.org/ 10. 1016/j.eswa.2024.121892.

[17] Liang, M., Yang, B., Zeng, W., et al. (2020). PnPNet: End-to-end perception and prediction with tracking in the loop. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11550–11559). IEEE. https:// doi. org/ 10.1109/CVPR42600.2020.01157.

[18] Pang, Z., Ramanan, D., Li, M., et al. (2023). Streaming motion forecasting for autonomous driving. In 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 7407–7414). IEEE. https://doi.org/ 10.1109/ IRO S55552.2023.10342110.

[19] Papais, S., Wang, L., Jain, M., Rezaei, B., & Waslander, S. L. (2026). BeyondSight: Object permanence for end-to-end autonomous driving. In European Conference on Computer Vision (ECCV). Springer.

[20] Lu, D., Du, H., Wu, Z., et al. (2025). Risk assessment in autonomous driving: A comprehensive survey of risk sources, methodologies, and system architectures. Autonomous Intelligent Systems, 5(1), 103–123. https://doi.org/10. 1007/ s43684-024-00072-9.

[21] Dal'Col, L., Oliveira, M., & Santos, V. (2025). Joint perception and prediction for autonomous driving: A survey. IEEE Transactions on Intelligent Transportation Systems, 26(12), 21427–21452. https://doi.org/10.1109/TITS.2025.3487124.

[22] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In International Conference on Machine Learning (ICML) (pp. 2130–2143). PMLR.

[23] Yang, P., Zheng, Y., Zhang, Q., et al. (2025). UncAD: Towards safe end-to-end autonomous driving via online map uncertainty. In 2025 IEEE International Conference on Robotics and Automation (ICRA) (pp. 6408–6415). IEEE. https://doi.org/ 10. 1109/ ICRA64210.2025.11009278.

[24] Caesar, H., Bankiti, V., Lang, A. H., et al. (2020). nuScenes: A multimodal dataset for autonomous driving. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 11618–11628). IEEE. https:// doi. org/ 10.1109/CVPR42600.2020.01164.

[25] Wei, Y., Zhao, L., Zheng, W., et al. (2023). SurroundOcc: Multi-camera 3D occupancy prediction for autonomous driving. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 21672–21683). IEEE. https://doi. org/ 10.1109/ICCV51070.2023.01984.

[26] Wang, X., Zhu, Z., Xu, W, et al. (2023). OpenOccupancy: A large scale benchmark for surrounding semantic occupancy perception. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 17804–17813). IEEE. https:// doi. org/10.1109/ICCV51070.2023.01638.

[27] Sun, P., Kretzschmar, H., Dotiwalla, X., et al. (2020). Scalability in perception for autonomous driving: Waymo Open Dataset. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2443–2451). IEEE. https://doi.org/10.1109/CVPR42600.2020.00252.

[28] Jiang, A., Sun, H., Zhao, H., et al. (2026). UniUncer: Unified dynamic static uncertainty for end to end driving [Preprint]. arXiv. https://arxiv.org/abs/2603.07686.

[29] Ivanovic, B., Lin, Y., Shrivastava, S., et al. (2022). Propagating state uncertainty through trajectory forecasting. In 2022 IEEE International Conference on Robotics and Automation (ICRA) (pp. 2351–2358). IEEE. https://doi.org/ 10. 1109/ ICRA 46639. 2022. 9811647.

[30] Deglurkar, S., Shen, H., Muthali, A., et al. (2024). System-level analysis of module uncertainty quantification in the autonomy pipeline. In 2024 IEEE 63rd Conference on Decision and Control (CDC) (pp. 1103–1109). IEEE. https://doi. org/ 10. 1109/ CDC54075.2024.10835946.

[31] Araujo, B., Teixeira, J. F., Fonseca, J., et al. (2024). The road to safety: A review of uncertainty and applications to autonomous driving perception. Entropy, 26(8), 634. https:// doi. org/10.3390/e26080634.

[32] Abdar, M., Pourpanah, F., Hussain, S., et al. (2021). A review of uncertainty quantification in deep learning: Techniques, applications and challenges. Information Fusion, 76, 243–297. https:// doi.org/10.1016/j.inffus.2021.06.008.

[33] Meding, I., Bodin, A., Tonderski, A., et al. (2023). You can have your ensemble and run it too — Deep ensembles spread over time. In 2023 IEEE/CVF International Conference on Computer Vision (ICCV) (pp. 4022–4031). IEEE. https://doi. org/ 10.1109/ICCV51070.2023.00374.

[34] Gao, J., Chen, M., Xiang, L., et al. (2026). A comprehensive survey on evidential deep learning and its applications. IEEE Transactions on Pattern Analysis and Machine Intelligence, 48(3), 2118–2138. https://doi.org/10. 1109/ TPAMI. 2025. 352 4112.

[35] Ryu, W., Yu, S., Moon, S., et al. (2025). SUPER-AD: Semantic uncertainty-aware planning for end-to-end robust autonomous driving [Preprint]. arXiv. https://arxiv.org/abs/2511.22865.