Intervene3D: Intervention-Based Controlled Inference for Multimodal Perception under Partial Observability

Constantino Msigwa[1] **Denis Bernard**[1] **Jaeseok Yun**[1,2*]

[1]Dept. of Future Convergence Technology [2]Dept. of Internet of Things · Soonchunhyang University, Asan, South Korea · *Corresponding author

NeurIPS 2026

Ubicomp Lab

Soonchunhyang University

Fusion Detectors Are Fully Amortized

1. Partial observability

Occlusion, LiDAR sparsity and camera–LiDAR drift leave many scenes consistent with the same measurements.

2. One forward pass

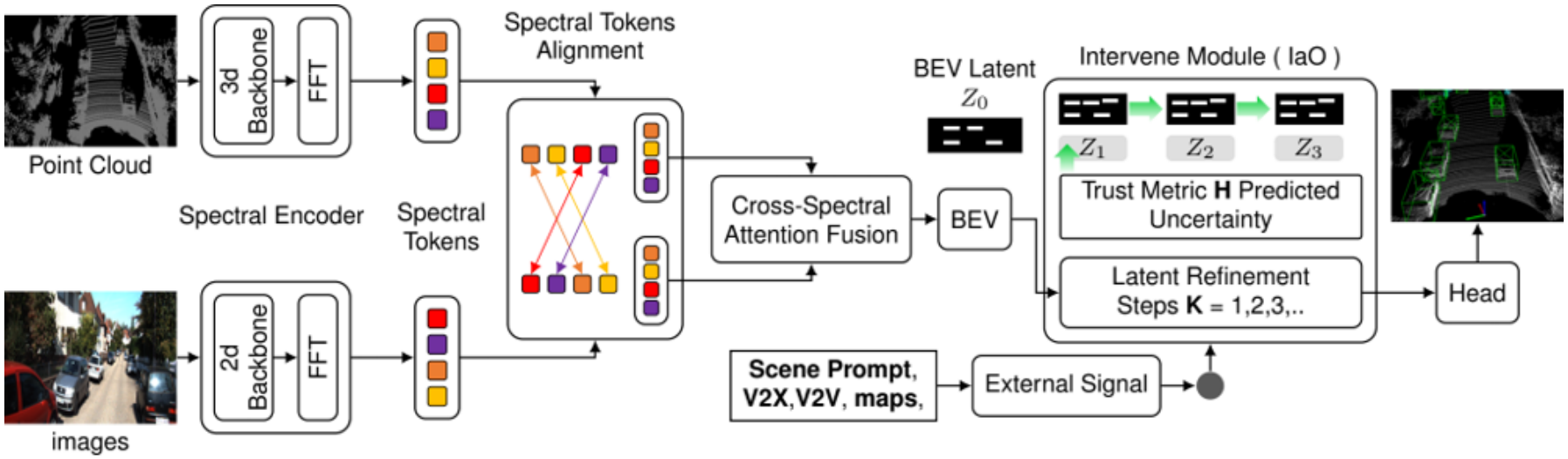
Evidence, uncertainty and priors are absorbed together. Injecting counts or regions entangles control with evidence.

$$Z = f(X, T)$$

3. What is missing

Enforce structured constraints, attenuate control when evidence is strong, and audit how much control changed the prediction.

Separate Evidence from Control



Feed-forward conditioning

$$Z = f(X, T)$$

Control mixed into the encoding.
Influence unbounded,
hard to audit.

Intervene3D (ours)

$$z_0 = \varphi(x), H(x)$$

Evidence encoded once.
Control $\tau \rightarrow$ energy $C(z;\tau)$.
Refinement bounded by H

Method: Three Components

1.**Spectral encoders**: Backbone → FFT → radial band tokens (image, LiDAR)

2. **Phase alignment**: Learned alignment, not exact calibration.

$$\Delta\varphi = f_{\Delta}(GAP(A_l), GAP(A_L)), \quad \varphi_L \leftarrow \varphi_L - \Delta\varphi$$

3. **IaO refinement**

- **Trust metric** $H(x) = \mathrm{diag}(\exp(-s(x)))$.
- **Control** $\tau = \Phi(x)$ is predicted from sensors
- (pretrained, then frozen and detached):

IaO: Bounded, Closed-Form Refinement

Constraint energy

$$C(z) = w(u) \sum_c [\alpha L_{\text{count}} + \beta L_{\text{spatial}}], \quad w(u) = \text{clip}(u^3, 0, 1)$$

- Count-interval loss
- Spatial-admissibility loss
- w(u): control fades when evidence is strong

Repeat K = 3 times

$$\tilde{z} = z_k - \eta H^{-1} \nabla C(z_k), \quad \|z - z_0\|_H^2 \leq 2\delta(u)$$

natural-gradient step — projection onto trust-region ellipsoid

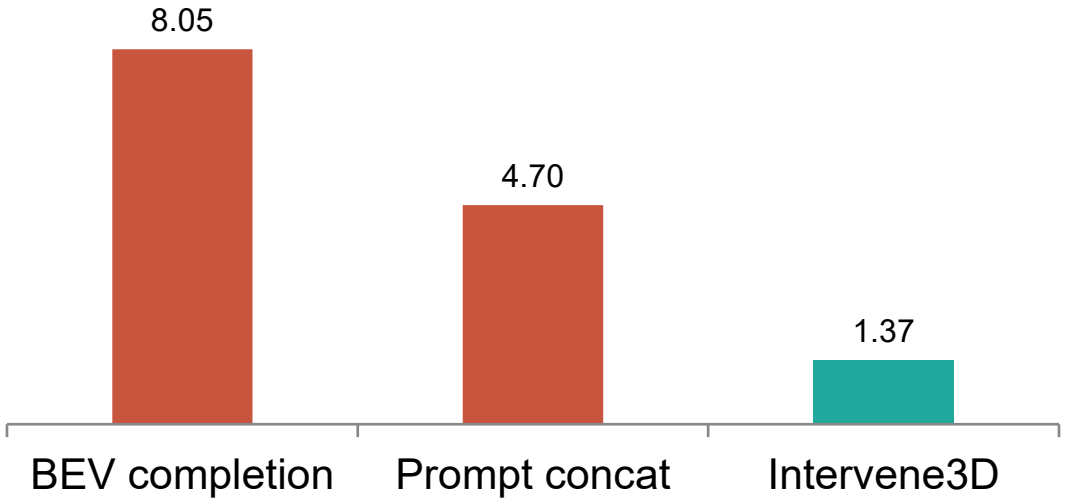
Detector fixed · Φ frozen · diagonal H = closed form

Conflicting Control: Intervene3D Is Less Sensitive

Table 3: Robustness to Conflicting Semantic Control on KITTI (Mod. AP_{3D}). BEV completion is our reimplementation.

Method	Normal	Conflict	ΔAP ↑	FP infl. ↓
BEV completion	82.90	74.85	-8.05	+20%
Prompt concat	88.36	83.66	-4.70	+12%
Intervene3D	**89.75**	**88.38**	**-1.37**	**+4%**

Moderate 3D AP drop (points, smaller is better)



False-positive inflation

+20% BEV completion

+12% Prompt concat

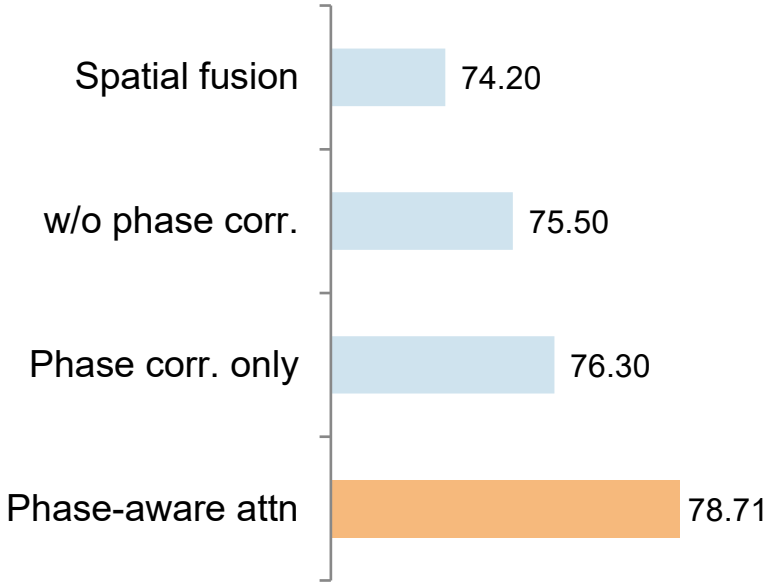
+4% Intervene3D

Robustness under Degraded Sensor Observations

Robustness-AUC

Method	Sparse	Extri.	Occl.
BEVFusion	78.20	70.15	88.40
Residual refiner	83.95	76.81	89.35
Intervene3D	83.83	78.71	89.45

Isolation study, Extri. AUC



Bars start at 72

Competitive Accuracy with Modest Overhead

Table 2: 3D mAP(%) results for the Car on the KITTI `test` set using R40.

Method	Easy	Moderate	Hard
F-PointNet [22]	82.19	69.59	60.59
3D-CVF [44]	89.20	80.05	73.11
Focals Conv [3]	90.55	82.28	77.59
SFD [34]	**91.73**	84.76	77.92
TED-M [32]	91.61	**85.28**	80.68
Intervene3D	90.84	84.62	**81.22**

0
trust-region violations

81.22
KITTI test Car Hard (best)

72.1 / 74.9
nuScenes mAP / NDS

KITTI val, Moderate AP3D

Class	w/o IaO	Intervene3D
Car	89.56	89.75
Pedestrian	59.79	62.59
Cyclist	67.88	68.14

Mean trust radius
0.84 → 0.93 (clean → occlusion)

Cost vs BEVFusion
110 vs 102 GFLOPs

Zero observed latent trust-region violations

Key Takeaways

Our Contributions

1. **Evidence Encoding:** Encode camera and LiDAR evidence once to build a reliable 3D scene representation.

2. **Controlled Inference:** Refine predictions using structured semantic constraints without re-encoding sensor evidence.

3. **Bounded Intervention:** Keep refinements within an uncertainty-adaptive, auditable trust region to limit changes to the original evidence.

Limitations & Future Work

1. **Control Generalization:** Main experiments use sensor-predicted control; external guidance such as HD maps and V2X remains unverified.

2. **Uncertainty Calibration:** Inaccurate uncertainty estimates may restrict useful corrections or allow excessive refinement.

3. **Latent-Space Guarantee:** Bounded refinement does not guarantee correct detections or prevent false positives.

Thank you

Code & configs:
github.com/Constantino2019-et/Intervene3D

Questions?

Constantino Msigwa
20237479@sch.ac.kr