Transcendental Idealism of Planner: Evaluating Perception from Planning Perspective for Autonomous Driving

Wei-Xin Li and Xiaodong Yang   QCraft, Inc.

QCRAFT

Introduction

- Problem
  - Real-world road tests for autonomous driving are costly.
  - How to evaluate **the impact of perception errors on an autonomous vehicle (AV)** offline?
- Baselines
  - Traditional metrics: nuScenes detection score (NDS) [1]
    - ☐ Ignore the response of an AV to errors.
  - AV-centric metrics: support distance error (SDE) [2]
    - ☐ Prior knowledge and handcrafted rules incoporated for metric design easily defeated by the problem complexity.
  - Result-centric metrics: planner KL divergence (PKL) [3]
    - ☐ Rely on weak correlation between the change in AV behaviour (the planning result) and the error consequence.

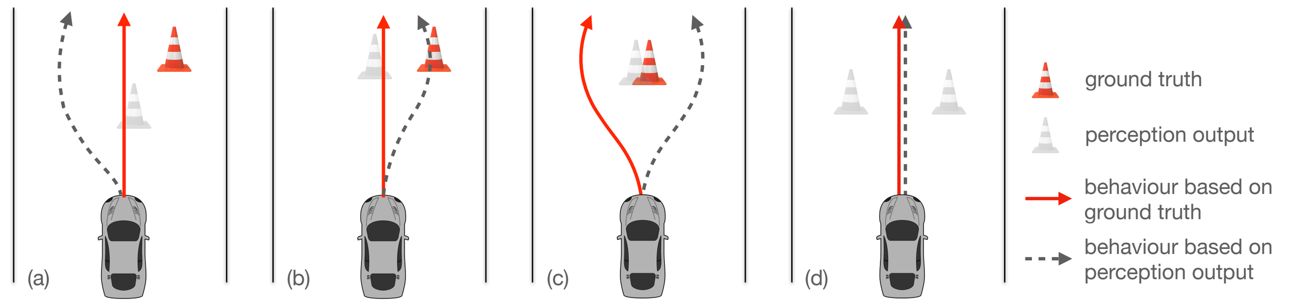

Figure 1: The change in AV behaviour due to a perception error is not always correlated to the consequence. The error consequence in (a) ('making a large detour') is far less significant than that in (b) ('hitting an object'), though the trajectory change in the former is greater. In (c) the consequence of either way is almost indifferent to the AV, yet the change in behaviour is considerable. In (d), two falsely detected cones are close to the AV on both sides when passing by without collision; the AV decides to move forward as in the ground truth case—the AV's behaviour remains unchanged regardless of the perception error.

- Our solution: **transcendental idealism of planner (TIP)**
  - Evaluate **the change in the planning process** due to a perception error to infer the consequence unbiasedly.

References

[1] Caesar, H., Bankiti, V., Lang, A., Vora, S., Liong, V. E., Xu, Q., Krishnan, A., Pan, Y., Baldan, G., and Beijbom, O. nuScenes: A multimodal dataset for autonomous driving. *CVPR*, 2020.

[2] Deng, B., Qi, C. R., Najibi, M., Funkhouser, T., Zhou, Y., and Anguelov, D. Revisiting 3D object detection from an egocentric perspective. *NeurIPS*, 2021.

[3] Philion, J., Kar, A., and Fidler, S. Learning to evaluate perception models using planner-centric metrics. In *CVPR*, 2020.

[4] Osborne, M. and Rubinstein, A. *A Course in Game Theory.* MIT Press, 1994.

[5] Zhu, B., Jiang, Z., Zhou, X., Li, Z., and Yu, G. Class-balanced grouping and sampling for point cloud 3d object detection. *arXiv preprint arXiv:1908.09492*, 2019.

Expected Utility Maximisation

- Planning as **expected utility maximisation (EUM)** [4]
  AV: an intelligent agent striving for the maximum reward:

$$a^* = \operatorname{argmax}_{a\in\mathcal{D}_a} EU(F_S, a),\ EU(F_S,a) := \mathbb{E}\left[U(S,a)\right]. \quad (1)$$

a^*: the optimal action given the ground truth perception;
U: the utility function (reward of doing a in state S);
$\mathcal{D}_a$: the set of all feasible AV actions;
$S \in \mathcal{S}$: the world state distributed as $F_S(s)$ in space $\mathcal{S}$.

- EUM in the Hilbert space

> Theorem 1 (Probability Measure Embeddings in $\mathcal{H}$) Let $\{\mathcal{X}, \mathtt{d}\}$ be a compact metric space with $\mathtt{d}$ as the metric, p be a Borel probability measure on $\mathcal{X}$, and X be a random variable on $\mathcal{X}$ with distribution function $F_X(x)$. If $F_X(x)$ is absolutely continuous with a square-integrable density function f_X ($f_X \in L^2$), then there exists a unique element $\mu_p \in \mathcal{H}$ such that
> $$\mathbb{E}_X\left[g(x)\right] = \langle \mu_p, g\rangle_{\mathcal{H}},\ \forall g \in \mathcal{H}, \quad (2)$$
> where element μ_p denotes the *embedding* of probability measure p in the Hilbert space $\mathcal{H} = (L^2, \langle\cdot,\cdot\rangle)$, with the inner product given by $\langle g,h\rangle_{\mathcal{H}} := \int_x g(x)h(x)\,\mathrm{d}x$.

With the embedding, the EUM of (1) can be rewritten as

$$a^* = \underset{a\in\mathcal{D}_a}{\operatorname{argmax}}\, \mathbb{E}_{p(s)}\left[U(s,a)\right] = \underset{a\in\mathcal{D}_a}{\operatorname{argmax}}\, \langle \mu_p, U_a\rangle_{\mathcal{H}}, \forall U_a \in \mathcal{H}.$$

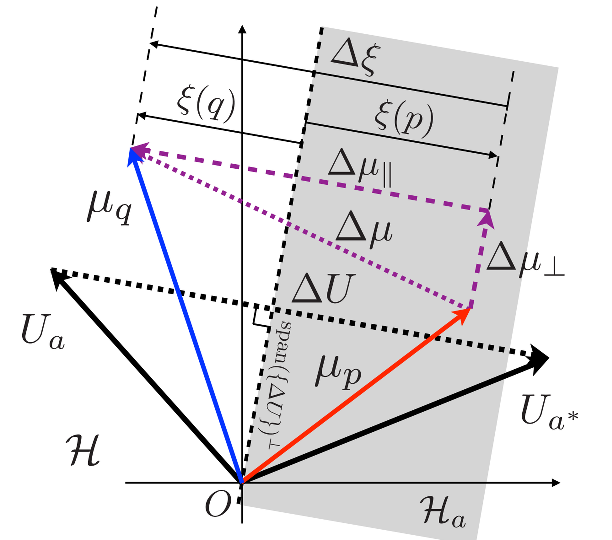

Figure 2: Illustration of EUM in $\mathcal{H}$.
p: ground truth (GT)
q: perception result
$\Delta U = U_{a^*} - U_a$: utility difference
$n_{\Delta U}$: behaviour direction
ξ: α^*-α preference score
$\Delta\xi$: change of α^*-α preference score
μ_p/μ_q: embedding of p/q
$\Delta\mu$: perception error
$\Delta\mu_\parallel$: planning-critical error (PCE)
$\Delta\mu_\perp$: planning-invariant error (PIE)

Efficient Unbiased Estimation

- Expected utilities are estimated by n empirical observations

$$EU_a = \frac{1}{n}\sum_{i=1}^n U(S_i, a),\ i.i.d.\ \{S_i\}_{i=1}^n \sim p_S. \quad (3)$$

> Theorem 2 (Exponential Convergence Rate) If there exists an $M \in \mathbb{R}$ such that $|U(S,a)| < M$ almost surely, then $\forall \varepsilon > 0$,
> $$\Pr\left(\left|EU_a - \mathbb{E}\left[U(S,a)\right]\right| > \varepsilon\right) < 2\exp\left(-\frac{n\varepsilon^2}{2M^2}\right).$$

- **Exemption** from the curse of dimensionality of space $\mathcal{S}$.
- **Flexibility** for $U(\cdot,a)$ and $p_S(\cdot)$ with any arbitrary forms.

Evaluating Perception via Planning

- Given $q(s)$, a^* is preferred over a by EUM if and only if
$$\xi(q; a^*, a) > 0 \quad (4)$$
with the α-β preference score given q ($\forall \alpha, \beta \in \mathcal{D}_a$)
$$\xi(q;\alpha,\beta) := \langle \mu_q, \Delta U(\alpha,\beta)\rangle = EU(q,\alpha) - EU(q,\beta). \quad (5)$$
- The a^*-a preference score (and the EUM result) may be affected by the perception error $\Delta\mu = \mu_q - \mu_p$:
$$\Delta\xi(a^*, a; q, p) := \xi(q; a^*, a) - \xi(p; a^*, a) = \langle \Delta\mu, \Delta U\rangle_{\mathcal{H}}.$$
- $\Delta\mu$ can be decomposed into two orthogonal components:
$$\Delta\mu = \Delta\mu_\parallel + \Delta\mu_\perp. \quad (6)$$
  - $\Delta\mu_\parallel$: the projection of $\Delta\mu$ onto behaviour direction $n_{\Delta U}$
  $$\Delta\mu_\parallel = \langle \Delta\mu, n_{\Delta U}\rangle_{\mathcal{H}}\, n_{\Delta U},\ n_{\Delta U} = \Delta U/\|\Delta U\|_{\mathcal{H}}. \quad (7)$$
  - $\Delta\mu_\perp \in \operatorname{span}(\{\Delta U\})^\perp$: the projection of $\Delta\mu$ onto the orthogonal complement of the subspace spanned by ΔU.
- **TIP score**: overall impact of a perception error on planning
$$\mathscr{I}(q,p;U,\mathcal{D}_a) := \min_{a\in\mathcal{D}_a} \Delta\xi(a^*, a; q, p) \leqslant 0. \quad (8)$$

Observations

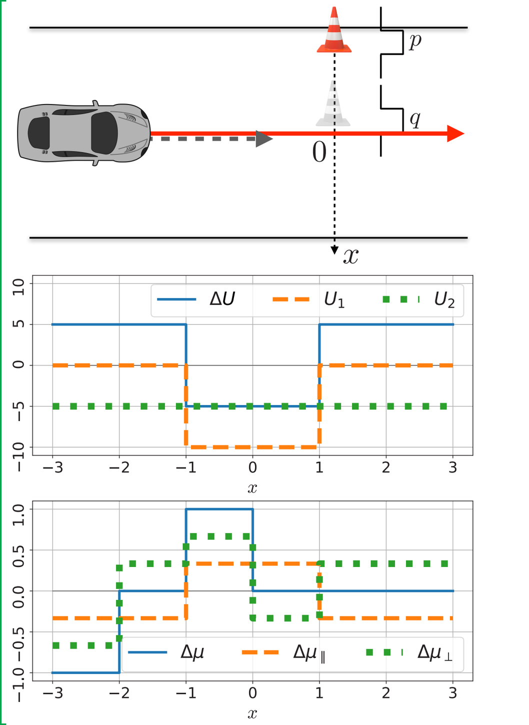

Figure 3: An example of PCE $\Delta\mu_\parallel$ and PIE $\Delta\mu_\perp$. An AV is moving forward on a 6m-wide road; a cone is in front on a line across the road (the x axis). The GT distribution of its location p is $\mathcal{U}_{[-3,-2]}$ (a uniform distribution over $[-3,-2]$); the perception result is $q = \mathcal{U}_{[-1,0]}$. The 2m-wide AV can either (i) keep moving forward (a^*, the solid arrowhead) with $U_1(x) = -10\cdot\mathbf{1}_{x\in[-1,1]}$ (x is the cone position); or (ii) hard brake to a full stop before the line (a, the dashed arrowhead) with $U_2(x) = -5$ (loss of hard braking is constant). In this example, PCE (PIE) accounts for 33% (67%) of the error energy.

- Perception errors affect planning diversely; subspace $\operatorname{span}(\{\Delta U\})^\perp$ contains all errors that do not interfere EUM.
- $\Delta\mu_\perp$ is denoted the **planning-invariant error (PIE)** since it does not affect EUM: $\langle \Delta\mu_\perp, \Delta U\rangle_{\mathcal{H}} = 0$.
- The change in a^*-a preference score is exclusively determined by $\Delta\mu_\parallel$, denoted the **planning-critical error (PCE)**:
$$\Delta\xi(a^*, a; q, p) = \langle \Delta\mu, \Delta U\rangle_{\mathcal{H}} = \langle \Delta\mu_\parallel, \Delta U\rangle_{\mathcal{H}}.$$
- Errors in subspace $\operatorname{span}(\{\Delta U\})$ either negatively affect planning ($\langle \Delta\mu, \Delta U\rangle < 0$), or favour a^* ($\langle \Delta\mu, \Delta U\rangle > 0$).

Empirical Study

- Basic settings
  - Production-grade module-based level-4 AV planners validated across megacities.
  - 1000 5-second scenarios (30-500 obejcts) with human annotation as GT for object detection.
- Results on synthetic data
  - A variery of sythetic noises are added to GT.
  - TIP renders better consistency in sensitivity and resolution.
  - Miss detection case study: TIP deems the one at the AV stopping distance (a barely avoidable one) most serious, as opposed to other baselines.

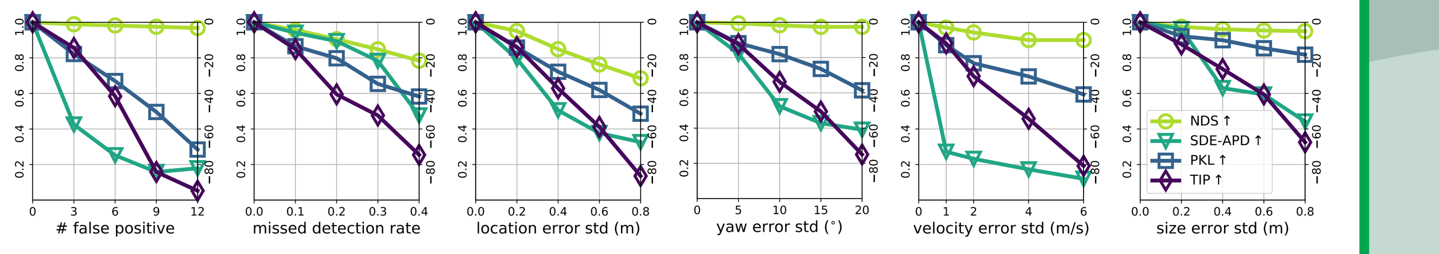

Figure 4: Metric comparison on different synthetic noises. The left (right) vertical axes are for NDS and SDE-APD (PKL and TIP).

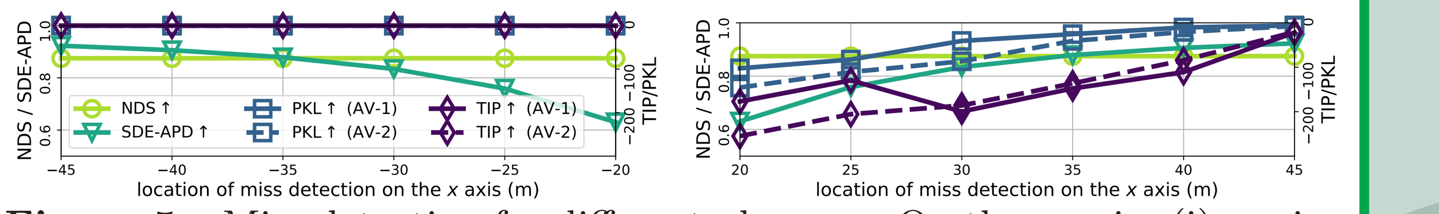

Figure 5: Miss detection for different planners. On the x-axis: (i) a miss detected stationary obstacle; (ii) a stationary vehicle (x=50m); (iii) an AV moving forward (x=0, 14m/s). AV-1(2): optimised for comfort (safety), braking capped at -4 (-6) m/s^2. The stopping distance is 30m (20m).

- Results on real data
  - Onboard 3D object detector: a pillar-based LiDAR network.
  - The optimal training ckeckpoint for planning is not the last one.
  - TIP finds critical errors that only cause minor behaviour changes (data points close to the x-axis in the scatter plot).

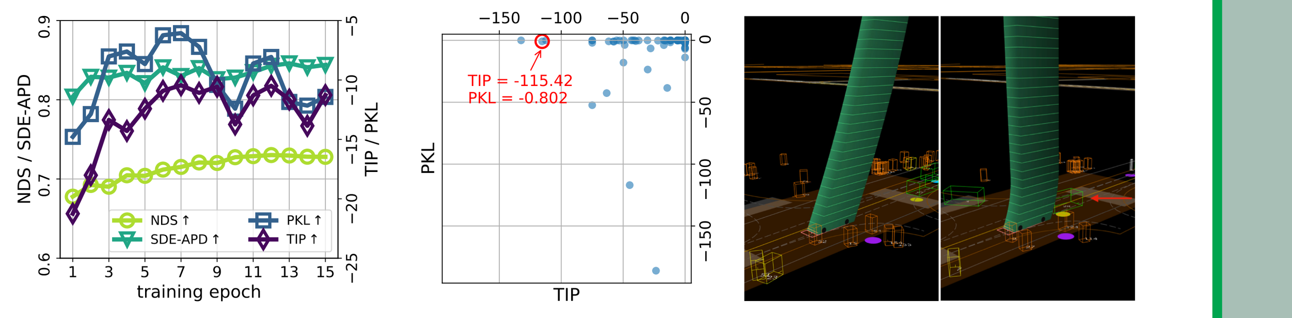

Figure 6: Metric comparison on real data. Left: metrics on different training checkpoints. Middle: scatter plot of TIP/PKL scores of different scenes. Right: the first one is a GT scene; the second one shows an outrageous false positive (highlighted by red arrows), which causes a jerk of -76.4m/s^3 (the typical limit is around -1.0m/s^3), despite a mild change in behaviour per PKL.

Application to Neural Planners

- The neural planner outputs AV probablistic locations [3].
- CBGS [5] detection results on the nuScenes dataset [1].
- TIP captures perception errors of mild consequences for planning though PKL considers them impactful.
- TIP finds that critical false positives (negatives) the planner is sensitive to are near the future AV path that require AV-object interaction.

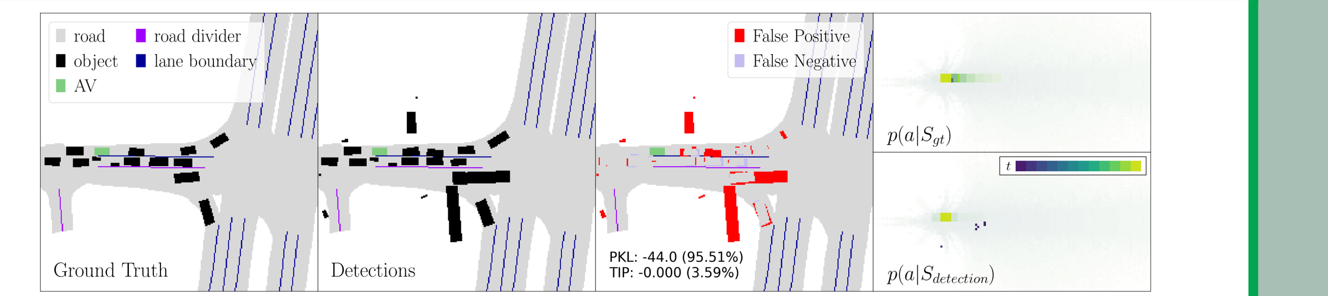

Figure 7: A scenario where PKL deems a large impact of the perception noise on planning yet TIP does not (score percentiles in the whole dataset are shown in parentheses). GT, the detection result, their difference, and planner outputs are shown in order.

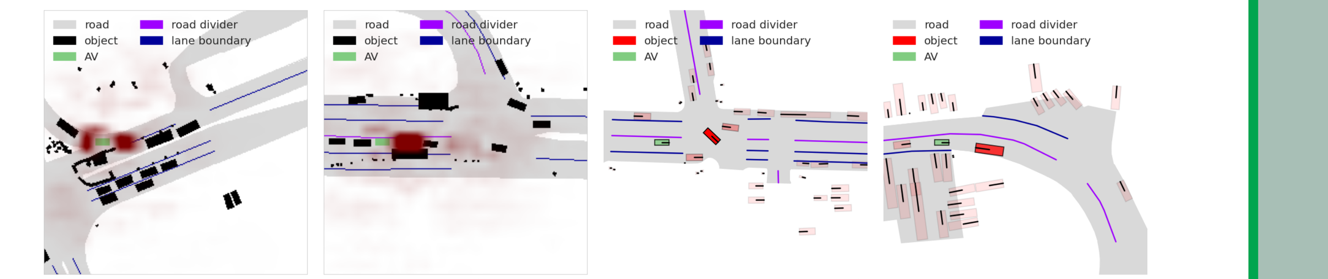

Figure 8: Significance (indicated by opacity) of false positives (left two) and negatives (right two) predicted by TIP.