Beyond Chamfer Distance

Granular, order-aware evaluation metrics for online mapping

Chouaib Bencheikh Lehocine Adam Lilja Junsheng Fu Lars Hammarstrand

Zenseact AB · Chalmers University of Technology

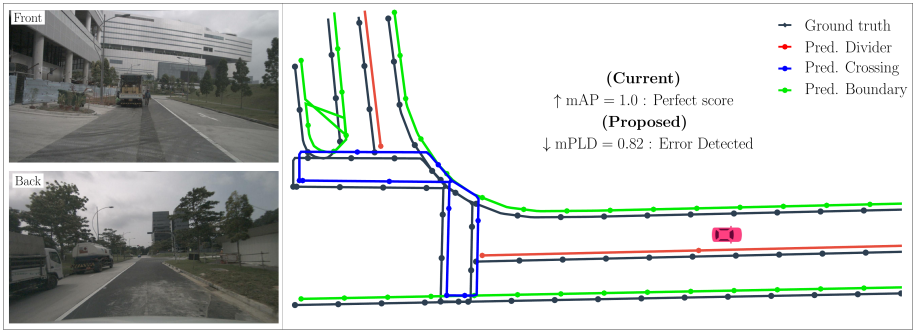
NeurIPS 2026 – Evaluations & Datasets

SOSPA – order-aware single-instance metric | **PLD** – soft multi-instance metric

The problem: Chamfer Distance + AP

Online map estimation (OME) predicts map elements as **ordered polylines/polygons**.

Evaluation uses **Chamfer Distance (CD) + Average Precision (AP)**



Predictions are diagonally offseted and one instance includes misordered points

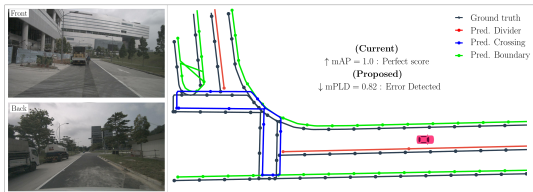
$\uparrow$ mAP $\in [0, 1] = 1.0$ (indicating perfect detection)

$\downarrow$ mPLD $\in [0, 1] = 0.82$ indicating degraded detection.

The problem: Chamfer Distance + AP

- CD **ignores point ordering** (polyline geometry)
- CD is coarse, outlier-sensitive, **not a metric**
- AP's hard TP/FP thresholds **discard positional accuracy**

Goal: an **order-aware**, **granular**, **metric** framework that jointly scores **detection *and* geometrical accuracy**



Predictions are diagonally offseted and one instance includes misordered points

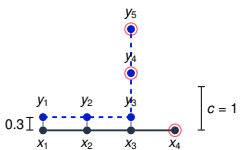
$\uparrow \text{mAP} \in [0, 1] = 1.0$ (indicating perfect detection)

$\downarrow \text{mPLD} \in [0, 1] = 0.82$ indicating degraded detection.

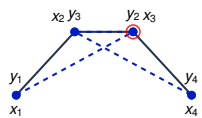
Pillar 1 — SOSPA: order-aware single-instance metric

SOSPA = *Sequence Optimal Sub-Pattern Assignment*

- adapts GOSPA from unordered sets to **ordered sequences** of points.
- SOSPA is a **true metric**



Partial match: 3 pairs matched, spurious points skipped at cost $c/2$ each.



Matched: (x_1, y_1), (x_2, y_3), (x_4, y_4); skipped x_3, y_2
SOSPA = 1.0 vs.CD = 0

Order error: same points, swapped order — SOSPA penalizes it; CD/GOSPA score 0.

Pillar 2 — PLD: soft multi-instance metric

PLD = *Polyline Localization and Detection*: instantiates **Probabilistic GOSPA** with normalized SOSPA as base metric.

$$d_{\mathrm{PLD}}^{(c,p)}(\mathbf{X},\mathbf{Y}) = \Big[\min_{\theta} \sum_{(i,j)\in\theta} \big(\min(r_i, r_j)\, \bar{d}_{\mathrm{SOSPA}}(x_i, y_j)^p + \tfrac{1}{2}|r_i - r_j| \big) + \tfrac{1}{2} \sum_{\text{unmatched}} r \Big]^{1/p}$$

$$\bar{d}_{\mathrm{PLD}} = \frac{2\, d_{\mathrm{PLD}}}{\left(\tfrac{1}{2}(R_{\mathbf{X}} + R_{\mathbf{Y}})\right)^{1/p} + d_{\mathrm{PLD}}} \in [0,1], \quad R_{\mathbf{X}} = \textstyle\sum_i r_i$$

- Replaces hard TP/FP thresholding with a **confidence-weighted soft assignment**
- Confidences $r_i \in [0,1]$;
- **metric** for $p = 1$;
- mean over classes $\Rightarrow$ **mPLD** (analogue of mAP)

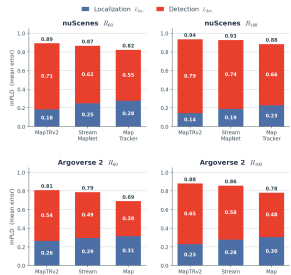
Interpretable decomposition ($p = 1$)

$$\bar{d}_{\mathrm{PLD}} = \underbrace{\bar{e}_{\mathrm{loc.}}}_{\text{TP geometry}} + \underbrace{\bar{e}_{\mathrm{det.}}}_{\substack{\text{misses+false alarms} \\ \text{+conf. mismatch}}} \in [0,1]$$

Benchmarks: PLD ranks methods *and* diagnoses them

Method	nuScenes ↑CD-AP	nuScenes ↓PLD	AV2 ↑CD-AP	AV2 ↓PLD
MapTRv2	0.183	0.936	0.492	0.879
StreamMapNet	0.219	0.928	0.512	0.856
MapTracker	**0.357**	**0.884**	**0.642**	**0.781**

$c = 1.5,\ p = 1.$



- mPLD **preserves the CD–mAP ranking**: MapTracker $>$ StreamMapNet $>$ MapTRv2.
- Detection error dominates localization by 2–4$\times$

Takeaways

- **SOSPA** — order-aware, granular, outliers-robust single-instance similarity
- **PLD** — soft multi-instance metric that scores **localization** and **detection**
- PLD preserves AP rankings
- PLD exposes that **detection is the current bottleneck** in online mapping

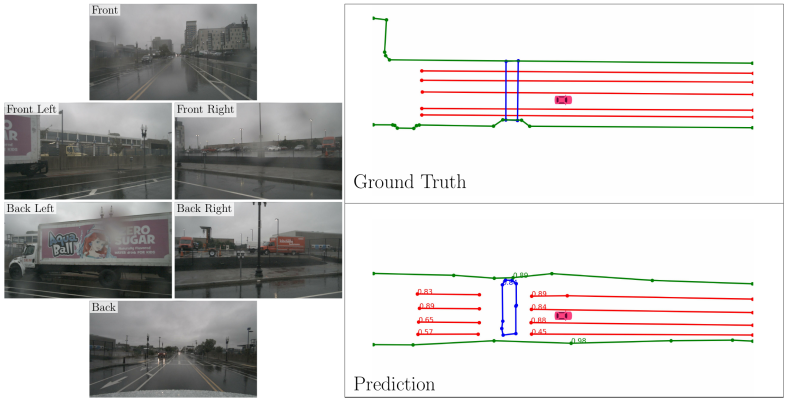
Code: `github.com/zenseact/SOSPA`

Paper: arXiv:2605.22578

arXiv:2605.22578

Additional Materials

Qualitative example: perfect AP, imperfect map



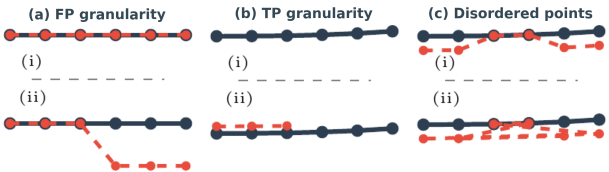
$\uparrow \mathrm{AP}_{\mathbf{Ped.}} = 1.0$ $\quad \downarrow \mathrm{PLD}_{\mathbf{Ped.}} = 0.824$ (Loc., Det.) $(0.658, 0.166)$

$\mathrm{AP}_{\mathbf{Div.}} = 0.0$ $\quad \mathrm{PLD}_{\mathbf{Div.}} = 0.869$ (Loc., Det.) $(0.193, 0.676)$

$\mathrm{AP}_{\mathbf{Bnd.}} = 1.0$ $\quad \mathrm{PLD}_{\mathbf{Bnd.}} = 0.722$ (Loc., Det.) $(0.610, 0.112)$

MapTracker on nuScenes R_{100}: $\uparrow$AP is 1.0 for crossings and boundaries, 0.0 for dividers; $\downarrow$PLD gives graded, per-class scores and **separates localization from detection error**.

Why it matters: toy cases CD/FD–AP cannot separate



Case	PLD	Loc.	Det.	CD-AP	FD-AP
(a) FP gran. (i)	**0.000**	0.000	0.000	1.0	1.0
(a) FP gran. (ii)	**0.800**	0.000	0.800	1.0	0.0
(b) TP gran. (i)	**1.000**	0.000	1.000	0.0	0.0
(b) TP gran. (ii)	**0.778**	0.222	0.556	0.0	0.0
(c) Disordered (i)	**0.593**	0.593	0.000	1.0	1.0
(c) Disordered (ii)	**0.778**	0.222	0.556	1.0	0.0

$c = 0.5,\ p = 1$; CD-/FD-AP use threshold 0.5.

- CD-AP (and often FD-AP) return **identical, uninformative** scores across all three cases.
- **PLD distinguishes** every case and attributes each error to **localization** vs. **detection**.

Long range and cost: FD–AP vanishes, PLD stays informative

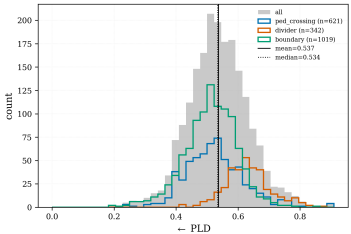
Method	↑ CD–mAP		↑ FD–mAP		↓ mPLD	
	R_{60}	R_{100}	R_{60}	R_{100}	R_{60}	R_{100}
MapTRv2	0.278	0.183	0.150	**0.029**	0.893	0.936
StreamMapNet	0.335	0.219	0.180	**0.031**	0.866	0.928
MapTracker	0.390	0.357	0.222	**0.091**	0.822	0.884

nuScenes; $c = 1.5$, $p = 1$; FD thresholds $\{1, 2, 3\}$ m at both ranges.

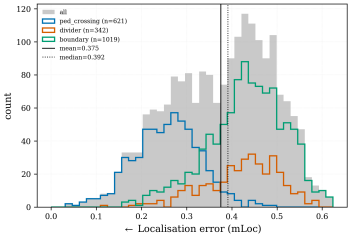
Evaluation time (StreamMapNet, nuScenes R_{60}, 64 workers)	CD–mAP	mPLD
Runtime (s)	38.0	37.8

- FD–mAP at R_{100} is ≈ 0.03–0.09: **uninformative** for comparing methods.
- mPLD keeps a clear spread and the same ranking at R_{100} (0.936 / 0.928 / 0.884).
- mPLD costs the same as CD–mAP (≈ 38 s).

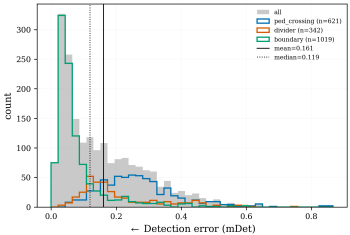
Near-perfect AP still hides errors (AP $>$ 0.95)



PLD (mean 0.54)



Localization



Detection

StreamMapNet predictions with AP $>$ 0.95 (nuScenes R_{60}): PLD is far from zero, and the residual error is mostly **detection** (confidence mismatch, misses, false alarms).

Pillar 1 — SOSPA: Equation

SOSPA = *Sequence Optimal Sub-Pattern Assignment*

- adapts GOSPA from unordered sets to **ordered sequences** of points.
- SOSPA is a **true metric**

$$d_{\mathrm{SOSPA}}^{(c,p)}(\mathbf{x},\mathbf{y}) = \Big(\min_{\theta \in \Theta^{\mathrm{ord}}} \sum_{(i,j)\in\theta} d(x_i,y_j)^p + \tfrac{c^p}{2}\,(n_x + n_y - 2|\theta|) \Big)^{1/p}$$

$$\bar{d}_{\mathrm{SOSPA}} = \frac{2\,d_{\mathrm{SOSPA}}}{\left(\tfrac{c^p}{2}(n_x+n_y)\right)^{1/p} + d_{\mathrm{SOSPA}}} \in [0,1]$$

- SOSPA is a **true metric**