Mapping like a Skeptic: Probabilistic BEV Projection for Online HD Mapping

Fatih Erdoğan[1] Merve Rabia Barın[1, 2] Fatma Güney[1, 2]
[1] Department of Computer Engineering, Koç University [2] Kuis AI Center

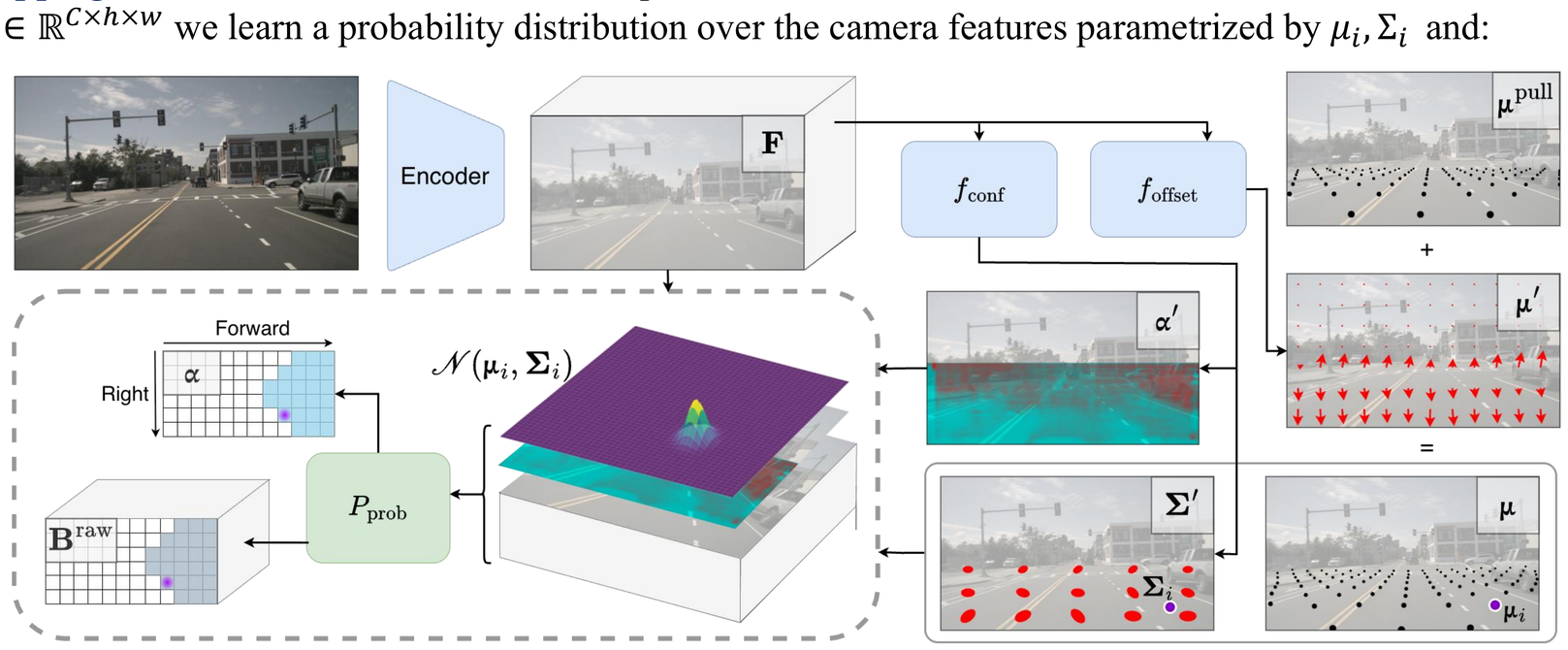
Introduction

Challenges: HD mapping is increasingly shifting toward vectorized HD mapping, where the map is represented as structured vector elements rather than dense rasters. Existing apporaches suffer from generalizability issues as it is evidenced by the the performance drop from old split to new split. We suspect that the culprit is the BEV feature extraction, as it is the building block for the vector module that generates the final road elements.

Proposal: A probabilistic BEV extraction method would better capture the road geometry and a confidence based temporal aggregation method would improve the quality of extracted features.

Methodology

Perspective to BEV Mapping: We first estimate a confidence map $\alpha' \in \mathbb{R}^{H \times W}$ over the camera features $\mathbf{F} \in \mathbb{R}^{C \times H \times W}$, and for each grid cell $\mathbf{b}_i \in \mathbb{R}^C$ in BEV feature map $\mathbf{B}^{\text{raw}} \in \mathbb{R}^{C \times h \times w}$ we learn a probability distribution over the camera features parametrized by μ_i, Σ_i and:



- sample K locations and compute their weights w_k such that:

$$w_k = \alpha'_k \cdot \mathcal{N}(\mu_{i,k};\, \mu_i, \Sigma_i) \quad \text{where} \quad \mu_{i,k} \sim \mathcal{N}(\mu_i, \Sigma_i), \quad \alpha'_k = P(\alpha', \mu_{i,k})$$

- compute $\mathbf{b}_i$ and corresponding confidence score α_i by:

$$\mathbf{b}_i = \sum_{k=1}^{K} w_k\, P(\mathbf{F}, \mu_{i,k}) \qquad \alpha_i = \sum_{k=1}^{K} w_k$$

with $P(\cdot, \cdot)$ denoting bilinear sampling and $\mathcal{N}(\mu_{i,k};\, \mu_i, \Sigma_i)$ being normalized over K gaussian likelihoods.

History Aggregation: We aggregate features extracted using the perspective features at $\mathbf{B}^{\text{raw}}$ with historical features $\mathbf{B}^{\text{raw}}_{\text{hist}}$ using their confidence maps and update the final confidence map as:

$$\mathbf{B}^{\text{raw}} = \alpha \odot \mathbf{B}^{\text{raw}} + \alpha_{\text{hist}} \odot \mathbf{B}^{\text{raw}}_{\text{hist}} \qquad \alpha = \alpha \odot \alpha + \alpha_{\text{hist}} \odot \alpha_{\text{hist}}$$

and normalize the results with pointwise summation of α and α_{hist}

Algorithm

Input:
- The set of cameras $\mathcal{C}$
- Historical raw BEV features $\mathbf{B}^{\text{raw}}_{\text{hist}}$
- Confidence scores α_{hist}
- Selected historical BEV features $\mathbf{B}_{\text{hist}}$
- Per-camera encoded features, distance masks, and pull indices $\{\mathbf{F}_c, \boldsymbol{d}_c^{\text{mask}}, \mu_c^{pull} \mid c \in \mathcal{C}\}$

Output:
- Predicted BEV features $\mathbf{B}$

for each camera $c \in \mathcal{C}$ **do**
- **Offset Prediction:** $\mu'_c \leftarrow f_{\text{offset}}(\mathbf{F}_c, \mathbf{d}_c^{\text{mask}})$
- **Cov. and Conf. Prediction:** $(\Sigma'_c, \alpha'_c) \leftarrow f_{\text{conf}}(\mathbf{F}_c)$
- **Offset Sampling:** $\mu_c^{\text{off}} \leftarrow P(\mu'_c, \mu_c^{pull})$
- **Update Pull Index:** $\mu_c \leftarrow \mu_c^{\text{pull}} + \mu_c^{\text{off}}$
- **Covariance Sampling:** $\Sigma_c \leftarrow P(\Sigma'_c, \mu_c)$
- **Probabilistic Projection:** $(\mathbf{B}_c^{\text{raw}}, \alpha_c) \leftarrow P_{\text{prob}}(\mathbf{F}_c, \mu_c, \Sigma_c, \alpha'_c)$

end for

Merge Camera Features: $\mathbf{B}^{\text{raw}}, \alpha \leftarrow M_{\text{avg}}(\{\mathbf{B}_c^{\text{raw}} \mid c \in \mathcal{C}\}), M_{\text{avg}}(\{\alpha_c \mid c \in \mathcal{C}\})$

Merge with Historical Raw Features: $(\mathbf{B}^{\text{raw}}, \alpha) \leftarrow M_{\text{conf}}(\mathbf{B}^{\text{raw}}, \boldsymbol{\alpha}, \mathbf{B}^{\text{raw}}_{\text{hist}}, \alpha_{\text{hist}})$

Parametric Temporal Fusion: $\mathbf{B} \leftarrow f_{\text{temporal}}(\mathbf{B}^{\text{raw}}, \mathbf{B}_{\text{hist}})$

Save $\mathbf{B}, \mathbf{B}^{\text{raw}}, \boldsymbol{\alpha}$

Return $\mathbf{B}$

Ablations

ID	$\mathcal{N}(\mu,\Sigma)$	α	$\mathbf{B}^{\text{raw}}_{\text{hist}}$	AP_p	AP_d	AP_b	mAP	$C\text{-}mAP$
A	✗	✗	✗	43.9	31.6	47.4	41.0	33.4
B_1	✓	✗	✗	48.8	35.2	46.6	43.5	34.9
B_2		✓		48.9	34.2	48.0	43.6	35.4
C_1	✗	✗	✓	44.2	32.6	46.0	41.0	33.2
C_2		✓		46.6	35.6	48.1	43.4	35.2
D	✓	✗	✓	47.4	34.2	48.6	43.4	35.3
E		✓		49.8	36.2	50.1	45.4	37.2

Quantitative Results

Range (m)	Dataset	Method	AP_p	AP_d	AP_b	mAP	$C\text{-}mAP$
60×30	nuScenes	StreamMapNet	31.6	28.1	40.7	33.5	22.2
		MapTracker	45.9	30.0	45.1	40.3	32.5
		Ours	49.8 (+8.5%)	36.2 (+20.7%)	50.1 (+11.1%)	45.4 (+12.7%)	37.2 (+14.5%)
	Argoverse2	StreamMapNet	61.8	68.2	63.2	64.4	54.4
		MapTracker	70.0	75.1	68.9	71.3	63.2
		Ours	72.9 (+4.1%)	76.9 (+2.4%)	67.5 (-2.0%)	72.4 (+1.5%)	62.7 (-0.8%)
100×50	nuScenes	StreamMapNet	25.1	18.9	25.0	23.0	14.6
		MapTracker	45.9	24.3	38.4	36.2	27.5
		Ours	52.5 (+14.4%)	33.1 (+36.2%)	42.9 (+11.7%)	42.8 (+18.2%)	33.5 (+21.8%)
	Argoverse2	StreamMapNet	60.1	56.1	47.5	54.6	41.3
		MapTracker	71.2	64.6	58.5	64.8	55.7
		Ours	74.9 (+5.2%)	67.4 (+4.3%)	58.9 (+0.7%)	67.1 (+3.5%)	56.1 (+0.7%)

Qualitative Results