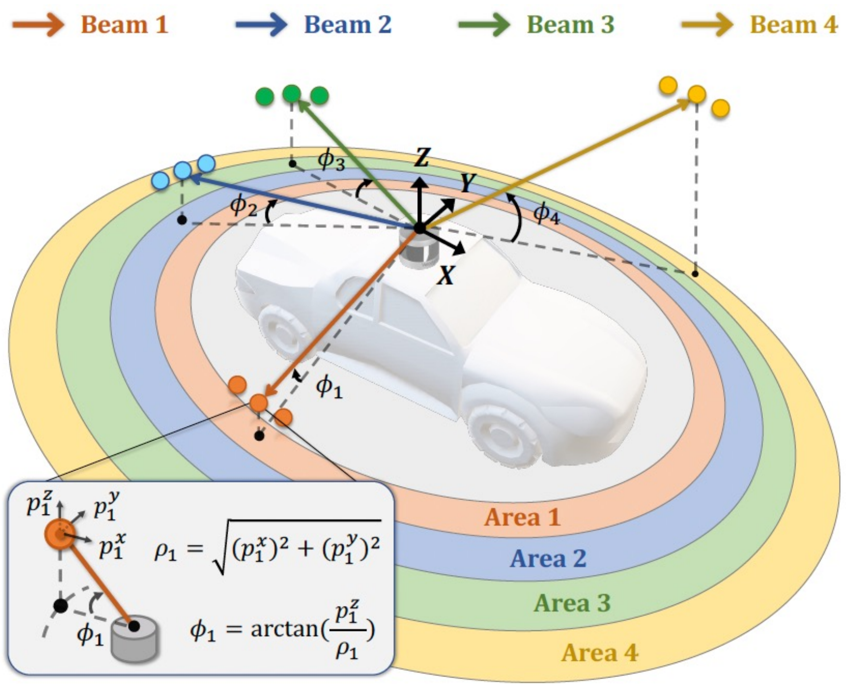
Derivation

LaserMix for Semi-Supervised LiDAR Semantic Segmentation

Patterns in LiDAR Scenes



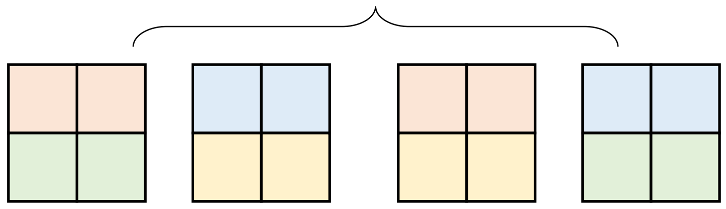
$H(X_{\text{in}}, Y_{\text{in}} \mid A = \text{Outer Rings})$

LiDAR data and labels strongly correlate with the area A, i.e.,

$H(X_{\text{in}}, Y_{\text{in}} | A)$ is low.

A Simplified Case:

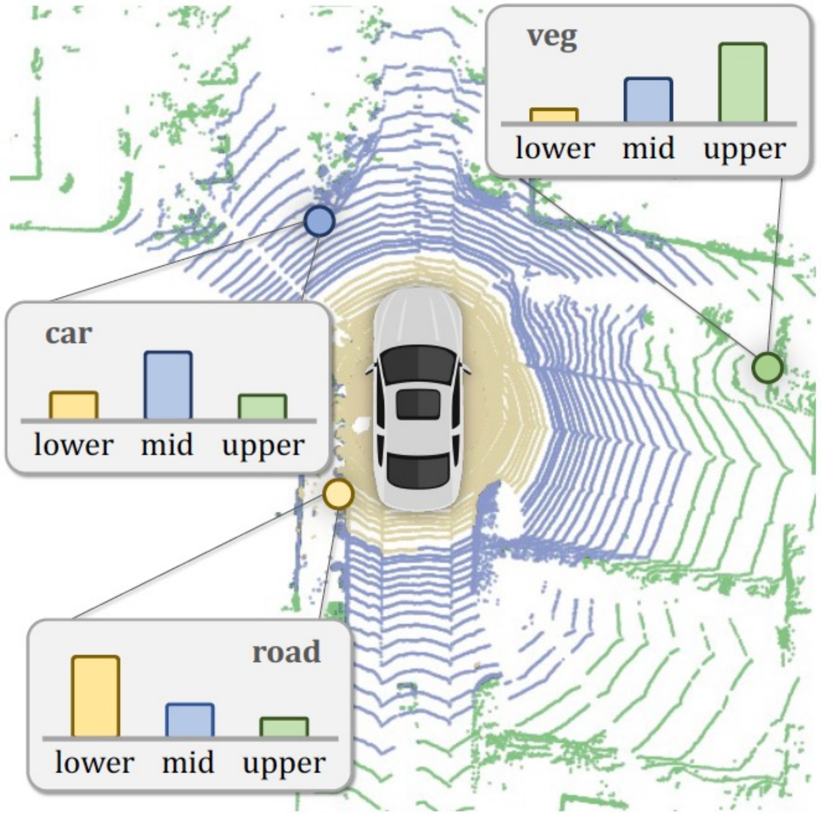
Color correlates with the row; each row has two colors



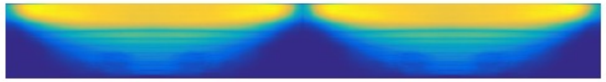
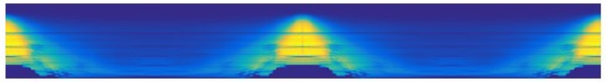
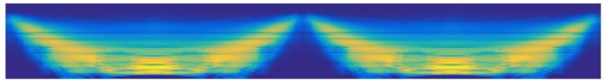
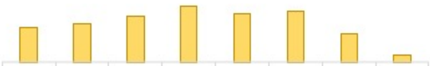
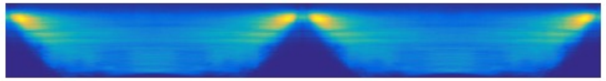
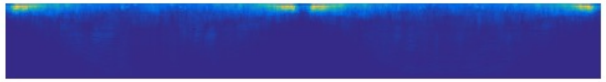
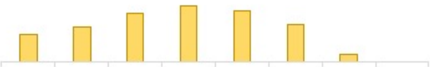
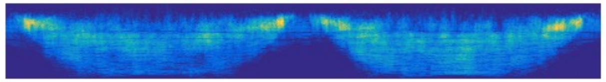
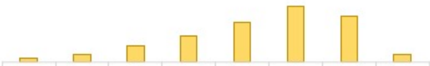
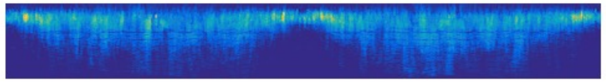
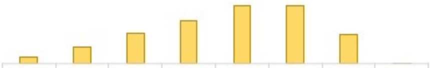
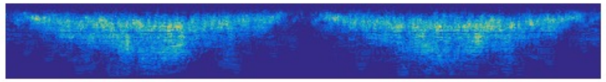
$$H(X, Y | A \in \{2\times2\}) = \log 4$$
$$H(X, Y | A \in \{1\times2, 1\times2\}) = \log 4$$
$$H(X, Y | A \in \{2\times1, 2\times1\}) = \log 2$$

Spatial Prior - Definition

- $H(X_{\text{in}}, Y_{\text{in}}|A)$ is low => $H(Y_{\text{in}}|X_{\text{in}}, A)$ is low (conditional entropy).
- Let θ be the parameter of the LiDAR segmentation network.
- We would like to solve the following:
  - $E_\theta[H_\theta(Y_{\text{in}}|X_{\text{in}}, A)] = c$, where c is a constant.
  - $\sum_\theta P(\theta) = 1$ (sum to one).
- Principle of Maximum Entropy:
  - $P(\theta) \propto \underbrace{\exp(-\lambda\, H_\theta(Y_{\text{in}}|X_{\text{in}}, A))}_{\text{spatial prior}}$, where λ is the Lagrange multiplier.



Spatial Prior - Evidence

Class	Type	Proportion	Distribution	Heatmap
vegetation	static	24.825%		
road	static	22.545%		
sidewalk	static	16.353%		
car	dynamic	4.657%		
traffic-sign	static	0.061%		
motorcycle	dynamic	0.045%		
person	dynamic	0.036%		
bicycle	dynamic	0.018%		

Statistics calculated from the SemanticKITTI dataset.

Certain class tends to appear at certain areas around the ego-vehicle!

Empirical Entropy By Marginalization

- $P(\theta) \propto \exp(-\lambda\, H_\theta(Y_{\text{in}}|X_{\text{in}}, A))$ => spatial prior.
- Compute the empirical entropy:
- $\hat{H}_\theta(Y_{\text{in}}|X_{\text{in}}, A) = E_{X_{\text{in}},Y_{\text{in}},A}[P_\theta(Y_{\text{in}}|X_{\text{in}}, A) \log P_\theta(Y_{\text{in}}|X_{\text{in}}, A)]$.
- $P_\theta(Y_{\text{in}}|X_{\text{in}}, A)$ means predicting the labels by the data inside an area A.
- The segmentation network predicts from full data. Therefore, we need X_{out} to complement the remaining area outside A and marginalize X_{out}.
- $P_\theta(y_{\text{in}}|x_{\text{in}}, a) = \frac{1}{|X_{\text{out}}|} \sum_{x_{\text{out}} \in X_{\text{out}}} P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})$.

MAP Estimation

- $P(\theta) \propto \exp(-\lambda\, H_\theta(Y_{\text{in}}|X_{\text{in}}, A))$ => spatial prior.
- $P_\theta(y_{\text{in}}|x_{\text{in}}, a) = E_{X_{\text{out}}}[P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})]$ => marginalization.
- We maximize the following posterior:
  - $C(\theta) = -\lambda \hat{E}_{x_{\text{in}} \in X_{\text{in}}, y_{\text{in}} \in Y_{\text{in}},\, a \in A}[H]$,
  - $H = P_\theta(y_{\text{in}}|x_{\text{in}}, a) \cdot \log P_\theta(y_{\text{in}}|x_{\text{in}}, a)$,
  - H is minimized only when $P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})$ is certain and consistent to x_{out}.

Concluding Remark

- $H = \frac{1}{|X_{\text{out}}|} \sum_{x_{\text{out}} \in X_{\text{out}}} P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}}) \log P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})$.
- $H = 0$ only when $P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})$ is certain and consistent to x_{out}.
- For every selected area and the data *inside* that area, a LiDAR segmentation network should make certain and consistent predictions regardless of the data *outside* the area.
- Directly compute $E_{y_{\text{in}} \in Y_{\text{in}}}[H]$ is infeasible / intractable, since $|y_{\text{in}}| = C^{H_{\text{in}} \times W_{\text{in}}}$ is too large.
- Instead, we use the pseudo-label to make sure that $P_\theta(y_{\text{in}}|x_{\text{in}}, a, x_{\text{out}})$ is certain and consistent.