Нейросетевые методы генерации HD-карт местности

Иванов Л.А.
ivanov.la@phystech.edu

Центр когнитивного моделирования МФТИ

2025 – Научно-исследовательский семинар



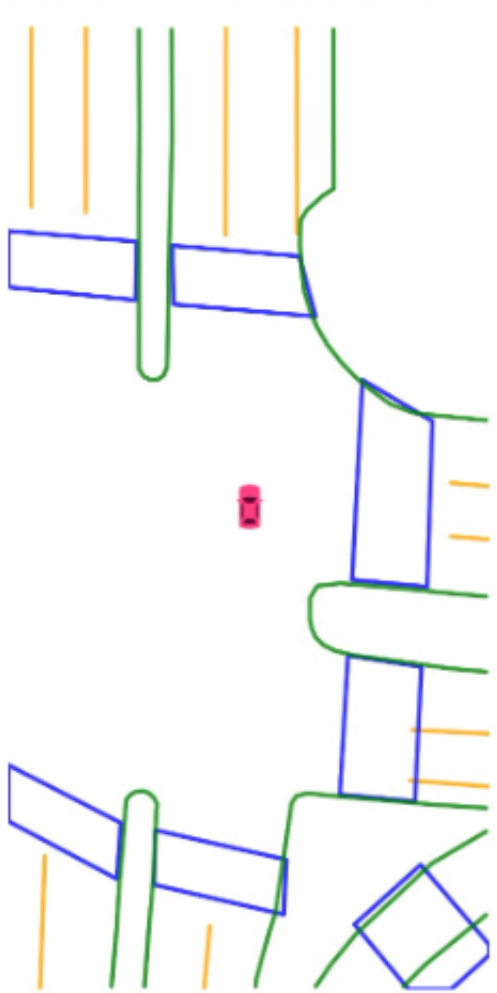


Мотивация

Карты высокой точности (HD Maps) являются одним из важнейших компонентов технологии автономного вождения при решении ряда задач:

- Навигации транспортного средства
- Прогнозировании поведения других участников движения
- Планировании собственного движения беспилотного транспорта

Цель работы Разработать эффективный алгоритм генерации HD-карты местности.



Метрики

- *Chamfer Distance* - метрика оценки схожести между двумя множествами точек, измеряющая среднее расстояние между точками одного множества и ближайшими точками другого множества.

$$\mathrm{CD}_{\mathrm{Dir}}(S_1, S_2) = \frac{1}{S_1} \sum_{x \in S_1} \min_{y \in S_2} \|x - y\|_2 \tag{1}$$

$$\mathrm{CD}(S_1, S_2) = \mathrm{CD}_{\mathrm{Dir}}(S_1, S_2) + \mathrm{CD}_{\mathrm{Dir}}(S_2, S_1) \tag{2}$$

- *AP (Average Precision)* — это метрика измерения того, насколько точно модель определяет и классифицирует элементы карты.

$$\mathrm{AP} = \frac{1}{10} \sum_{r \in \{0.1, 0.2, \ldots, 1.0\}} \mathrm{AP}_r \tag{3}$$

Каждое предсказание классифицируется как истинно положительное (TP) или ложноположительное (FP) в зависимости от значения *Chamfer Distance*.

Постановка задачи

Входные данные: $D_t = (d_t^1, d_t^2, ..., d_t^N)$, где d_t^i — данные, полученные от i-го сенсора (лидара, камеры, GPS и др.) в момент времени t.
Выходные данные: карта $M_t = \{(P^1, c^1), (P^2, c^2), \ldots, (P^k, c^k)\}$, где каждая полилиния P^i задаётся как последовательность координат $P^i = \{p_1^i, p_2^i, \ldots, p_{n_i}^i\}$, $p_j^i \in \mathbb{R}^2$, а c^i — метка класса.
Нужно: решить следующую задачу оптимизации

$$\text{AP}(M_t, F_\theta(D_t)) \to \max,$$

где AP — метрика Average Precision, а F_θ — метод генерации HD-карты с настраиваемыми параметрами θ.

Сравнение SOTA методов

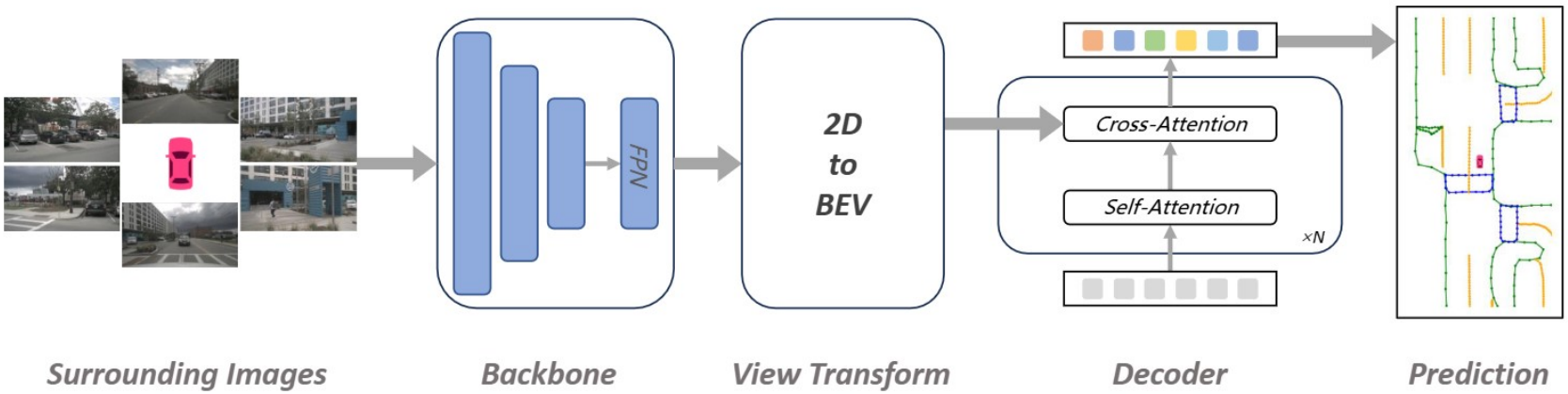
Center for Cognitive Modeling

Был выполнен запуск SOTA подходов генерации HD-карт.

Метод	Эпохи	mAP	Датасеты	Год
MapTR	24	50.0	nuScenes	2022
MapTR	110	59.3	nuScenes	2022
MapTRv2	24	61.3	nuScenes	2023
MapQR	24	66.3	nuScenes	2024
MapQR	110	72.6	nuScenes	2024

MapQR, 2024

В MapQR[1] — это трансформерный метод генерации HD-карт, который решает проблему отсутствия учета контекстной информации при генерации точек. Данный подход устраняет проблему случайной инициализации точек, обеспечивая их привязку к конкретному местоположению.



[1] Zihao Liu и др. (2024). «Leveraging Enhanced Queries of Point Sets for Vectorized Map Construction». В: *European Conference on Computer Vision*.

MapQR, 2024

В MapQR происходит моделирование на уровне сущностей (Instance Query), а не точек (Point Queries), как в MapTR[2].

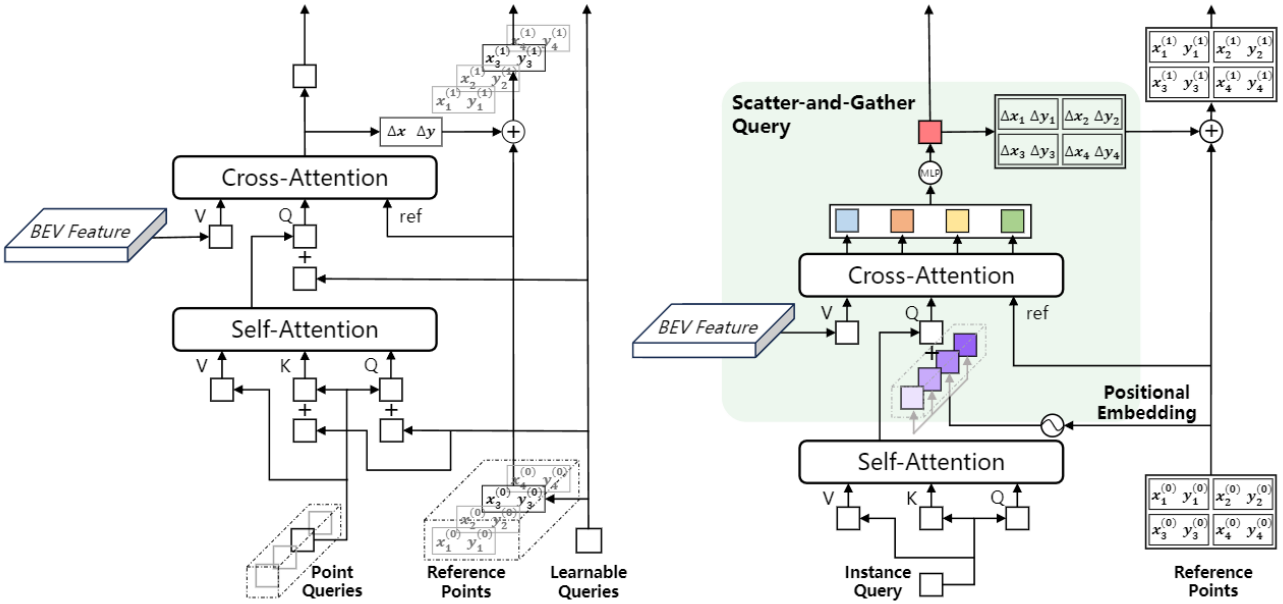


Рис.: Сравнение работы декодеров: **Слева** - MapTR. **Справа** - MapQR

[2] Bencheng Liao и др. (2022). «MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction». B.

Результаты работы MapQR

Center for Cognitive Modeling

При генерации элементов карты возникают проблемы с дрожанием и смещением точек.



Планируемые эксперименты

Гибридная функция потерь

.

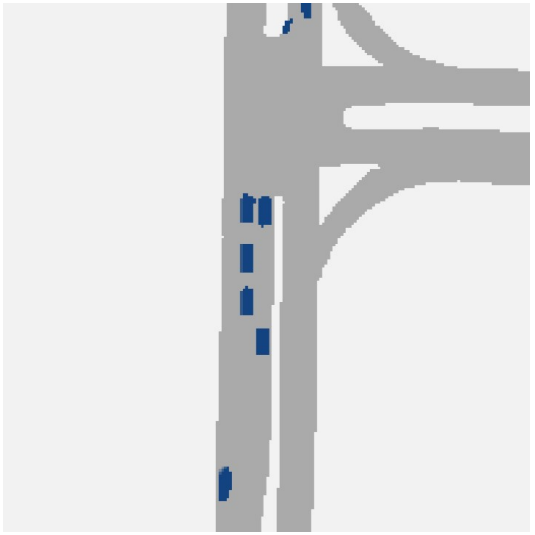
$$\mathcal{L}_{total} = \gamma_1 \mathcal{L}_{vec} + \gamma_2 \mathcal{L}_{dist} + \gamma_3 \sum_{i=\Omega_{line}} \mathcal{L}_i^{len} + \gamma_4 \sum_{i=\Omega_{closed}} \mathcal{L}_i^{area}, \qquad (4)$$

где

- $\mathcal{L}_{vec}$ — ошибка векторного направления (Vector Direction Difference Loss, VDDL).
- $\mathcal{L}_{dist}$ — ошибка соответствия точек между двумя наборами (Chamfer Distance, CD).
- $\mathcal{L}_i^{len}$ — ошибка длины (Length Difference Loss, LDL).
- $\mathcal{L}_i^{area}$ — ошибка площади (Area Difference Loss, ADL).
- Ω_{line} — множество открытые кривых (разграничительные линии, границы дорог).
- Ω_{closed} — множество замкнутых кривых (пешеходные переходы).

Семантическая сегментация

Дополнительная сегментационная голова, которая будет выполнять семантическую сегментацию BEV представления окружающей среды.



Дополнительно

Center for Cognitive Modeling

Также планируется выполнить следующие эксперименты:

- Заменить backbone для извлечения признаков из изображений на более современный feature extractor (CLIP, DINOv2, Swin Transformer).
- Добавить модальность облака точек, то есть добавить еще один backbone для получения BEV признаков из облака точек.
- Добавить сегментацию облака точек для добавления семантической информации о среде в BEV представлении, используемом для генерации HD-карты.

Спасибо за внимание!

ivanov.la@phystech.edu