Нейросетевые методы генерации HD-карт местности

Иванов Л.А.
ivanov.la@phystech.edu

Центр когнитивного моделирования МФТИ

2024 – Научно-исследовательский семинар





Мотивация

Center for Cognitive Modeling

Карты высокой точности (HD Maps) являются одним из важнейших компонентов технологии автономного вождения. Использование информации об окружающей среде, такой как границы полос движения, разделительные полосы и пешеходные переходы, играет решающую роль при решении задач:

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

Цель работы Разработать эффективный алгоритм генерации HD-карты местности.

Поставновка 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: $M_{1:t} = \{M_1, M_2, ..., M_t\}$, где M_t включает все пространственные и семантические данные, характеризующие окружающую среду в момент времени t.
Нужно: решить задачу оптимизации следующего вида

$$\hat{M}_{1:t} = \arg\max_{M_{1:t}} P(M_{1:t}|D_{1:t}),$$

где $\hat{M}_{1:t}$ - наиболее вероятного состояния карты.

Классификация методов

Нейросетевые методы построения HD-карт можно классифицировать по нескольким различным критериям:

- *по способу обработки данных* — **унимодальные** или **мультимодальные** методы
- *по способу декодирования данных* — **сверточный декодер** или **трансформерный декодер**
- *по использованию временной информации* — методы, использующие временную информацию, отличаются от статичных тем, что способны учитывать изменения в окружающей среде, например, движение транспортных средств или пешеходов

Сегментационные методы

HDMapNet, 2021

HDMapNet[1] первый обучаемый метод построения HD-карт. Мультимодальный подход выполняющий сегментацию *BEV* (Bird's Eye View) представления сцены.

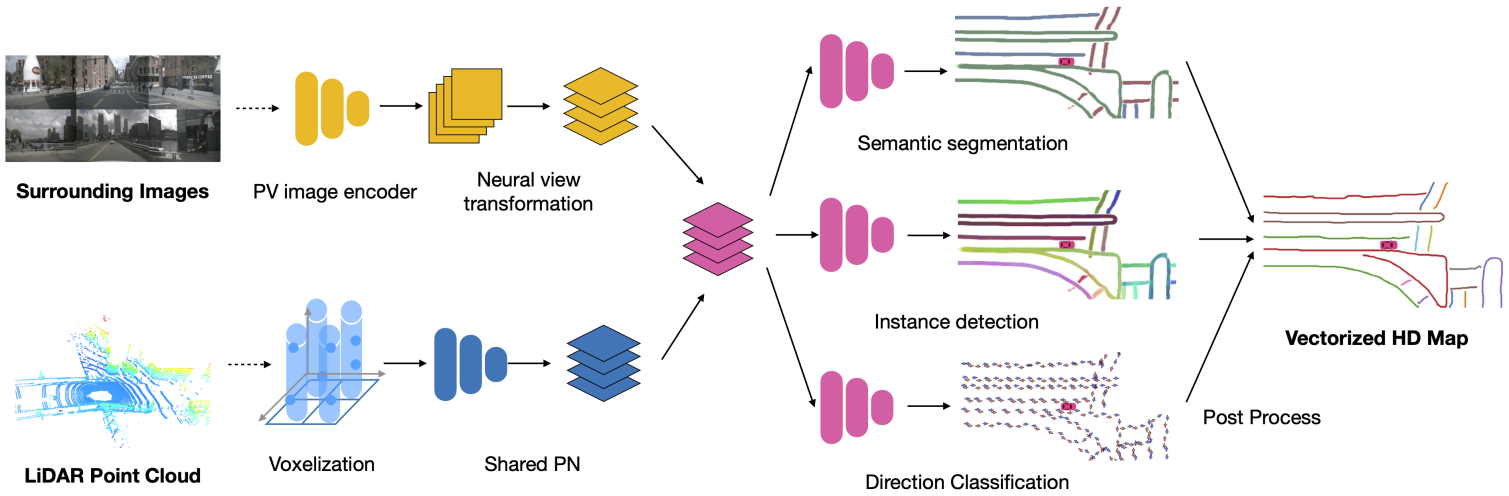

Рис.: Принцип работы алгоритма HDMapNet

[1] Qi Li и др. (2021). *HDMapNet: An Online HD Map Construction and Evaluation Framework*. arXiv: 2107.06307 [cs.CV].

Lidar2Map, 2023

Во время обучения Lidar2Map[2] выполняется дисстиляция знаний от кодировщика изображений к лидарной ветви, как на уровне признаков, так и на уровне вывода.

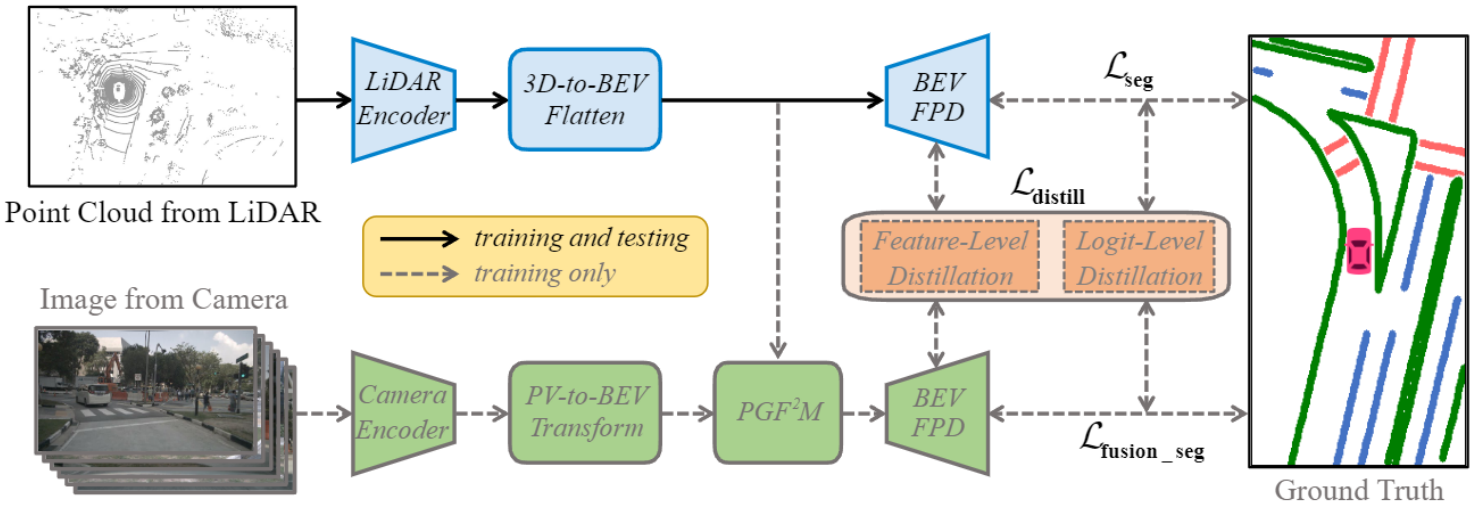

Рис.: Принцип работы алгоритма Lidar2Map

[2] Song Wang и др. (2023). «LiDAR2Map: In Defense of LiDAR-Based Semantic Map Construction Using Online Camera Distillation». В: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, с. 5186—5195.

Lidar2Map, 2023

Общая функция потерь при обучении выглядит следующим образом:

$$\mathcal{L} = \mathcal{L}_{seg} + \mathcal{L}_{fusion} + \mathcal{L}_{distill}, \tag{1}$$

$\mathcal{L}_{\textbf{distill}}$ включает в себя $\mathcal{L}_{\textbf{feature}}$ и $\mathcal{L}_{\textbf{logit}}$ для дистилляции знаний от ветви камеры к лидарной ветви

$$\mathcal{L}_{\textbf{feature}} = \sum_{i=1}^{N} \left\| \mathbf{M}^{\text{BEV}}_{\text{Fusion},i} - \mathbf{M}^{\text{BEV}}_{\text{LiDAR},i} \right\|_1 \tag{2}$$

$$\mathcal{L}_{\textbf{logit}} = \boldsymbol{D}_{\text{KL}} \left(\mathbf{P}^{\text{BEV}}_{\text{Fusion}} \| \mathbf{P}^{\text{BEV}}_{\text{LiDAR}} \right) \tag{3}$$

Трансформерные методы

VectorMapNet, 2022

VectorMapNet[3] первая трансформерная модель, которая для генерации элементов карты использует стратегию на основе авторегресси.

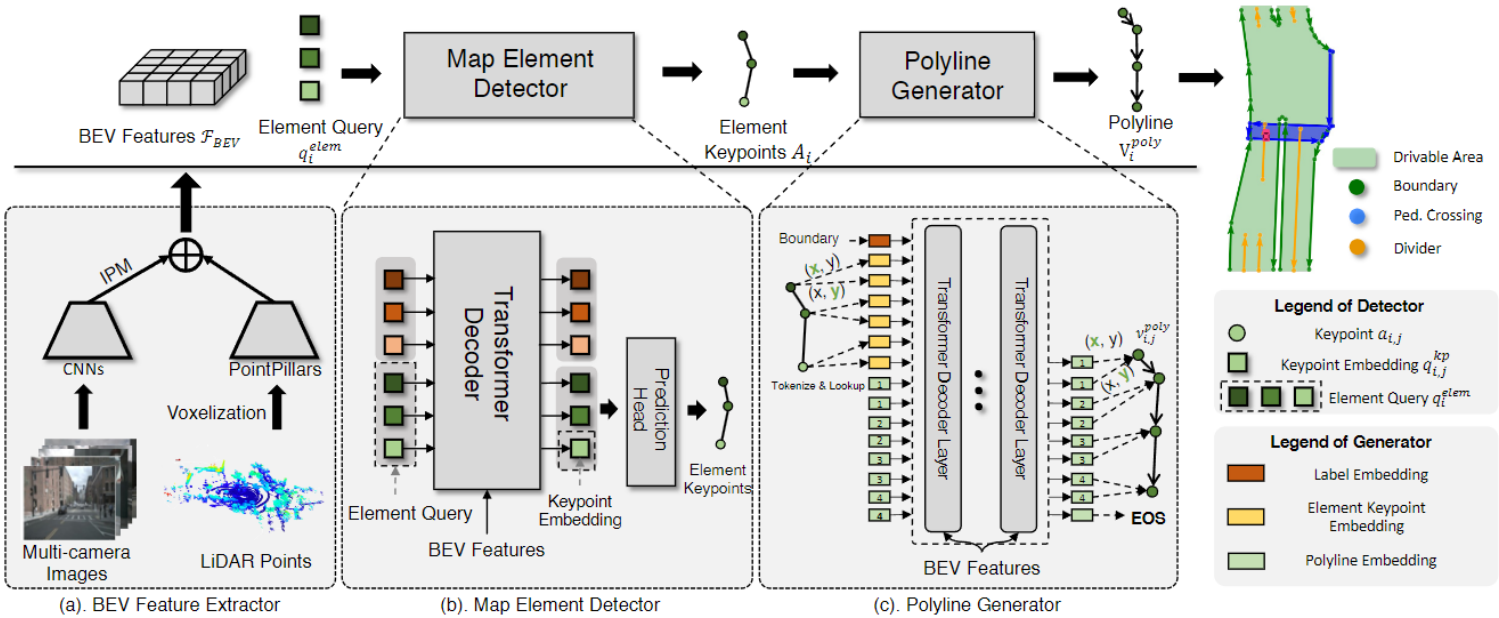

Рис.: Принцип работы алгоритма VectorMapNet

[3] Yicheng Liu и др. (2023). «VectorMapNet: End-to-end Vectorized HD Map Learning». В: *ICML*, с. 22352—22369.

MapQR, 2024

В MapQR[4] происходит моделирование на уровне сущностей (Instance Query), а не точек (Point Queries), как в MapTR[5].

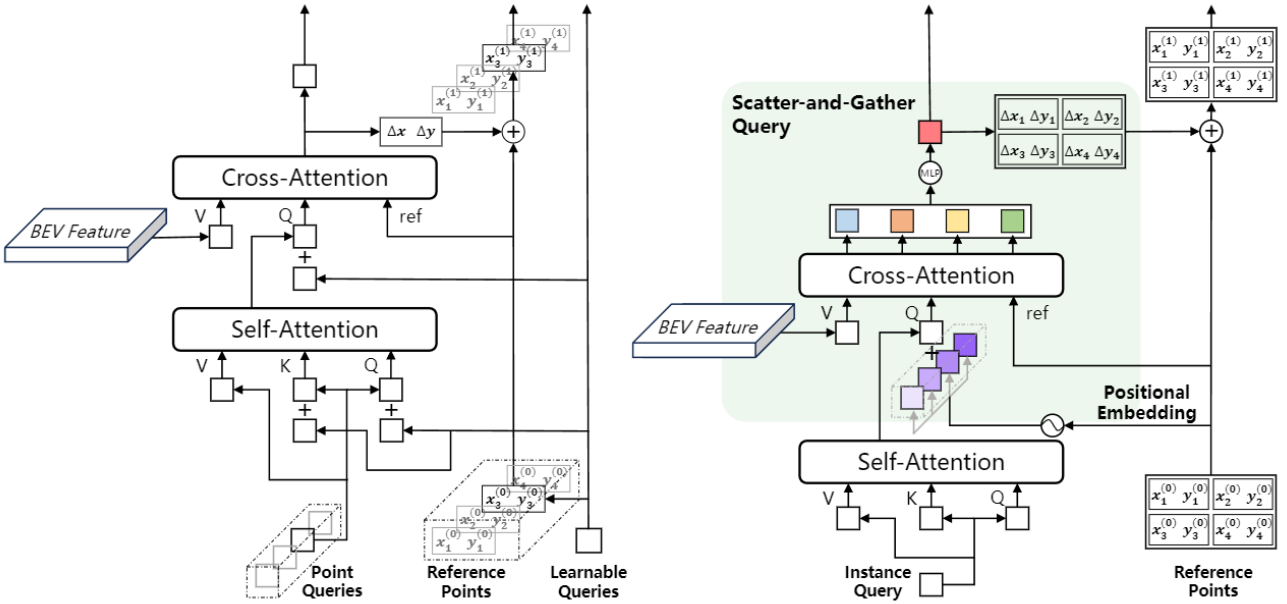

Рис.: Сравнение работы декодеров: **Слева** - MapTR. **Справа** - MapQR

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

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

Методы с использованием временной информации

StreamMapNet, 2023

StreamMapNet[6] использует буфер памяти, который хранит сгенерированные скрытые состояния, включающие в себя BEV признаки и обновленные запросы объектов.

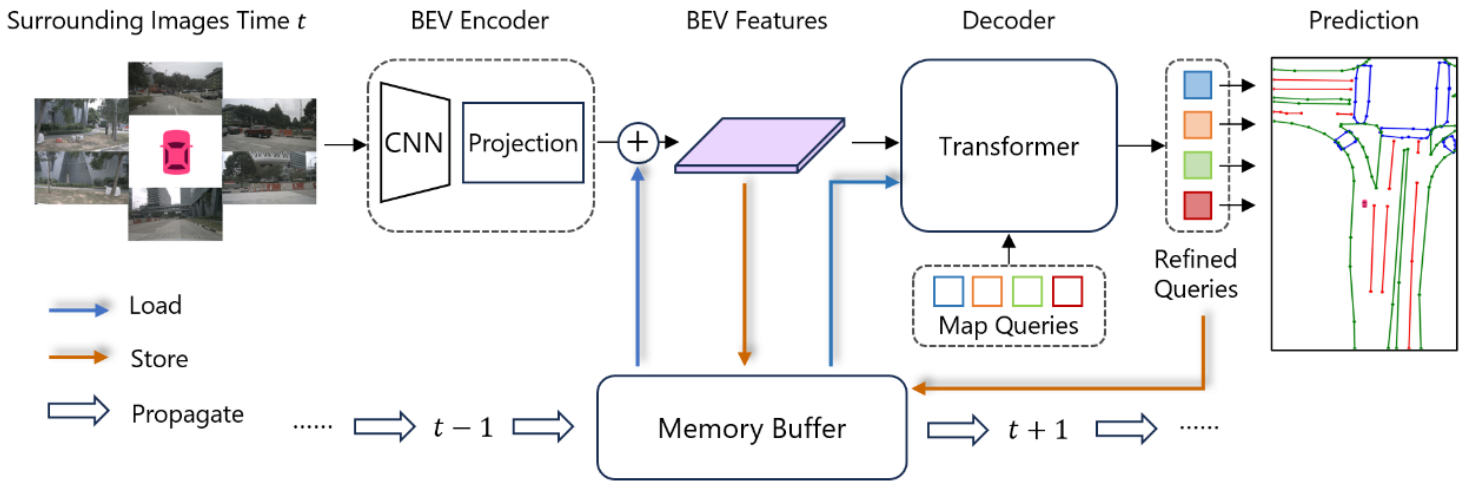

Рис.: Принцип работы алгоритма StreamMapNet

[6] Tianyuan Yuan и др. (2023). *StreamMapNet: Streaming Mapping Network for Vectorized Online HD Map Construction*. arXiv: 2308.12570 [cs.CV]. URL: https://arxiv.org/abs/2308.12570.

Таблица сравнения алгоритмов

Center for Cognitive Modeling

Method	Modality	Temporal info	Loc. Agnostic	Real-time	End-to-End	Year
HDMapNet	L+C	-	+	-	-	2021
SuperFusion	L+C	-	+	-	-	2022
VectorMapNet	L+C	-	+	-	+	2023
Lidar2Map	L	-	+	-	-	2023
MapTR	L+C	-	+	+	+	2023
StreamMapNet	C	+	-	+	+	2023
ADMap	L+C	-	+	+	+	2024
MapQR	C	-	+	+	+	2024
MapTracker	C	+	-	-	+	2024

Заключение

- Изначальные методы семантической сегментации с постобработкой имеют явное ограничение по скорости работы и качеству результата.
- Трансформерные модели значительно улучшили качество генерации HD-карт.
- Использования памяти позволяет повысить качество результатов, но требовательно к ресурсам и точности локализации.
- Для достижения наилучших результатов в генерации HD-карт необходимо использовать архитектуры трансформеров с учетом данных разных модальностей и временной информации.

Спасибо за внимание!

ivanov.la@phystech.ru

@LIv4noff