Разработка и исследование методов мультимодальной сегментации сцены

Матыкина О.В.
matykina.ov@phystech.edu

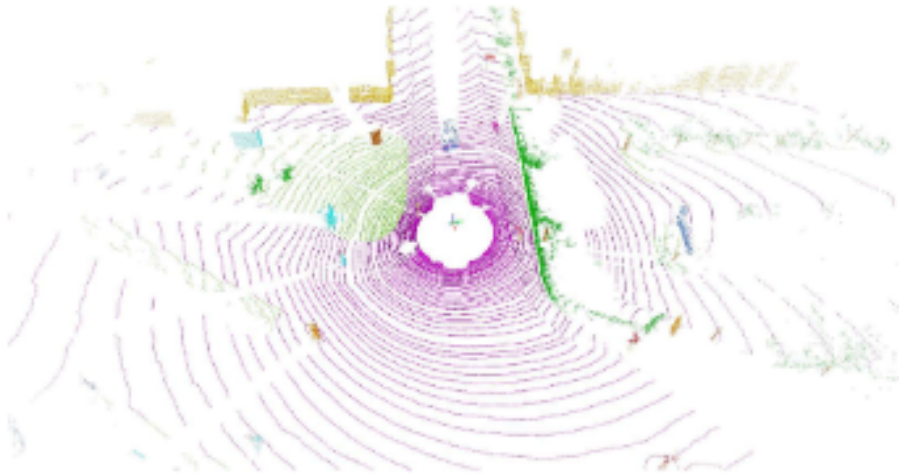
Магистерская программа
«Методы и технологии искусственного интеллекта»

Презентация статуса работы

Введение

Семантическая сегментация сцены - это мощный инструмент, который способствует развитию автономных транспортных средств, робототехники и других областей.
Результатом семантической сегментации является маска, которая показывает, какие области данных относятся к определенным классам объектов.





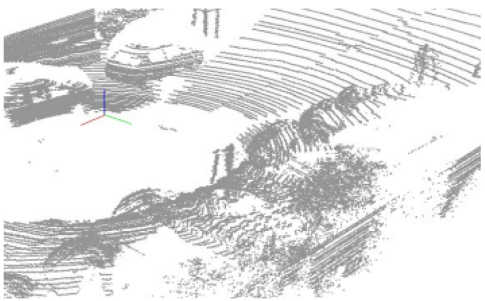
Пример семантической сегментации изображения и облака точек

Мотивация

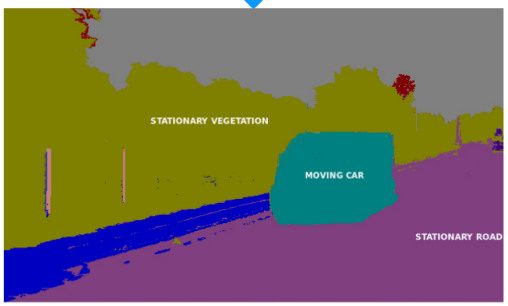
изображение с камеры



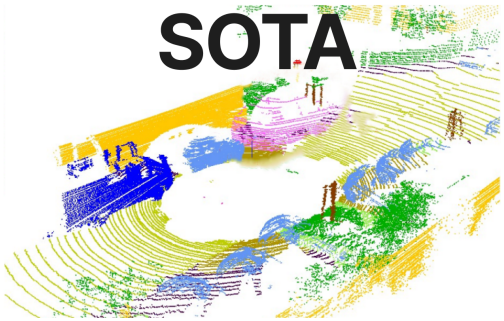
облако точек

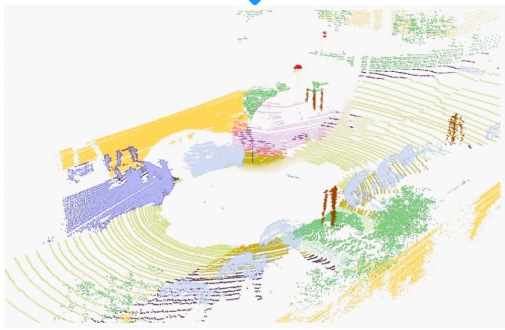


? слияние ?



2D сегментация





3D сегментация

Мотивация

Table 1
The influence level of various weather conditions on sensors.

Modality	Light rain <4 mm/h	Heavy rain >25 mm/h	Dense smoke /Mist vis<0.1 km	Fog vis<0.5 km	Haze/Smog vis>2 km	Snow	Strong light
LiDAR (λ 850–950 nm and 1550 nm)	2	3	5	4	1	5	2
Radar (24, 77 and 122 GHz)	0	1	2	0	0	2	0
Ground-Penetrating Radar (100–400 MHz)	0	0	0	0	0	1	0
Camera	3	4	5	4	3	2 (dynamic) 3 (static)	5
Stereo camera	Almost same as regular camera						

1

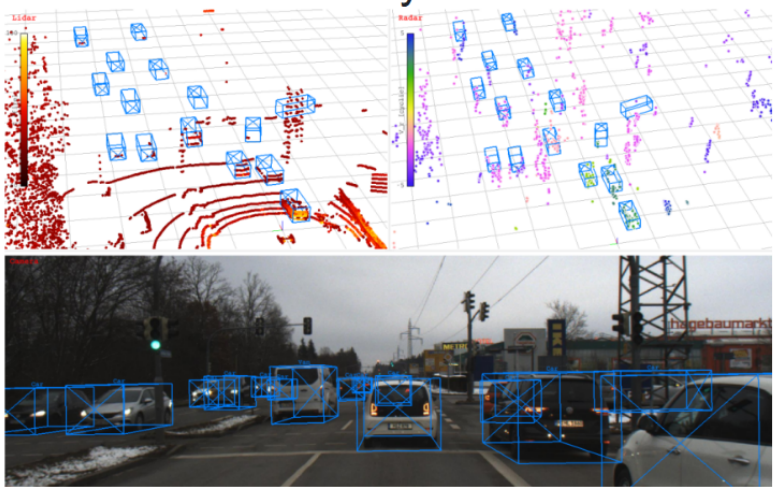
[1] Yuxiao Zhang и др. «Perception and sensing for autonomous vehicles under adverse weather conditions: A survey». В: *ISPRS Journal of Photogrammetry and Remote Sensing* 196 (2023), с. 146—177.

Постановка задачи

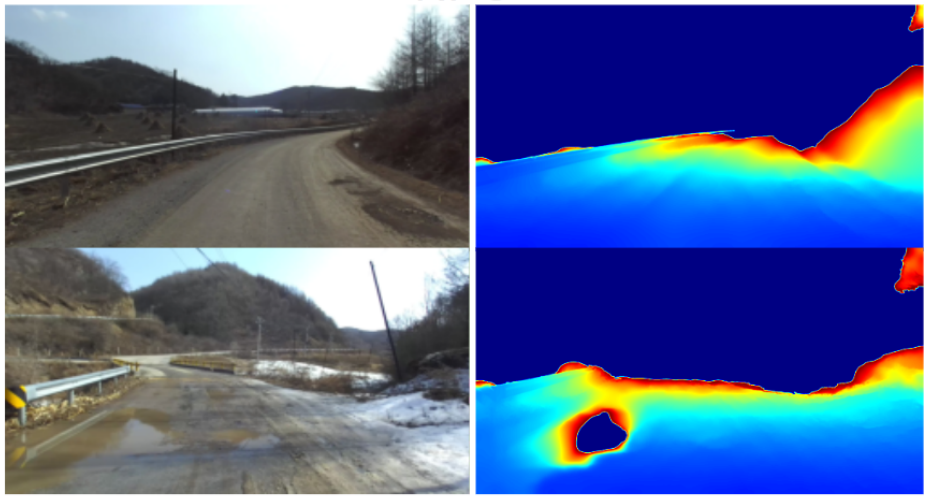
- На вход подается $\mathbf{R} = \{\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N\}$ - трехмерная сцена с облаком точек от радара, где каждая точка $\mathbf{r}_i$ имеет координаты (x_i, y_i, z_i), $\mathbf{D} = \{\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_M\}$ - трехмерная сцена с облаком точек, полученное из карты глубины Depth, где каждая точка $\mathbf{d}_j$ имеет координаты (x_j, y_j, z_j). Пусть $\mathbf{I} = \{\mathbf{i}_1, \mathbf{i}_2, \ldots, \mathbf{i}_M\}$ - изображение размерности $M = W_i \times H_i$.
- Требуется найти отображение $f : \mathbf{D} \times \mathbf{I} \to \mathbf{C}$, где $\mathbf{C}$ - множество классов, такое что для пары $(\mathbf{d}_j, \mathbf{i}_j)$ выполняется $f(\mathbf{d}_j, \mathbf{i}_j) = c$, где c - метка класса, соответствующая точке $\mathbf{d}_j$ и пикселю $\mathbf{i}_j$ с учетом облака точек радара $\mathbf{R}$.
- Оценка качества найденного отображения оценивается по метрике $\textit{Intersection over Union} = \dfrac{||A \cap B||}{||A \cup B||}$, где A и B - маски сегментации. При этом маски могут быть как на плоскости, так и в пространстве.

Мультимодальные датасеты

Astyx



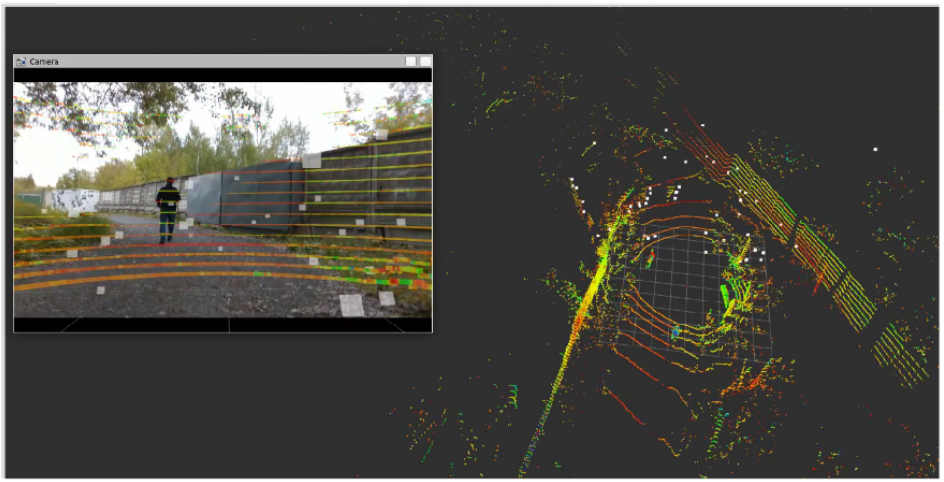
NPO



View-of-Delft



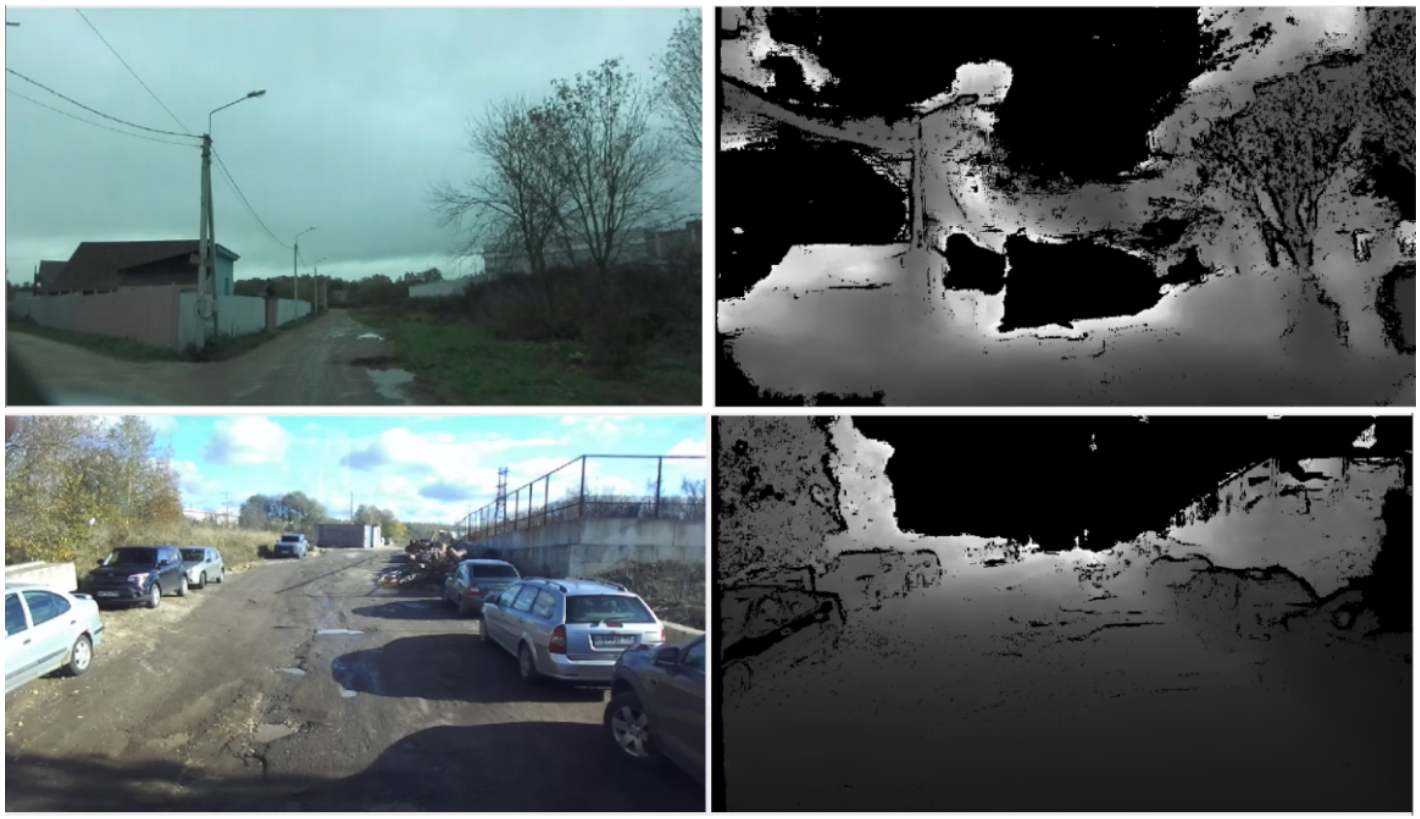
Campus



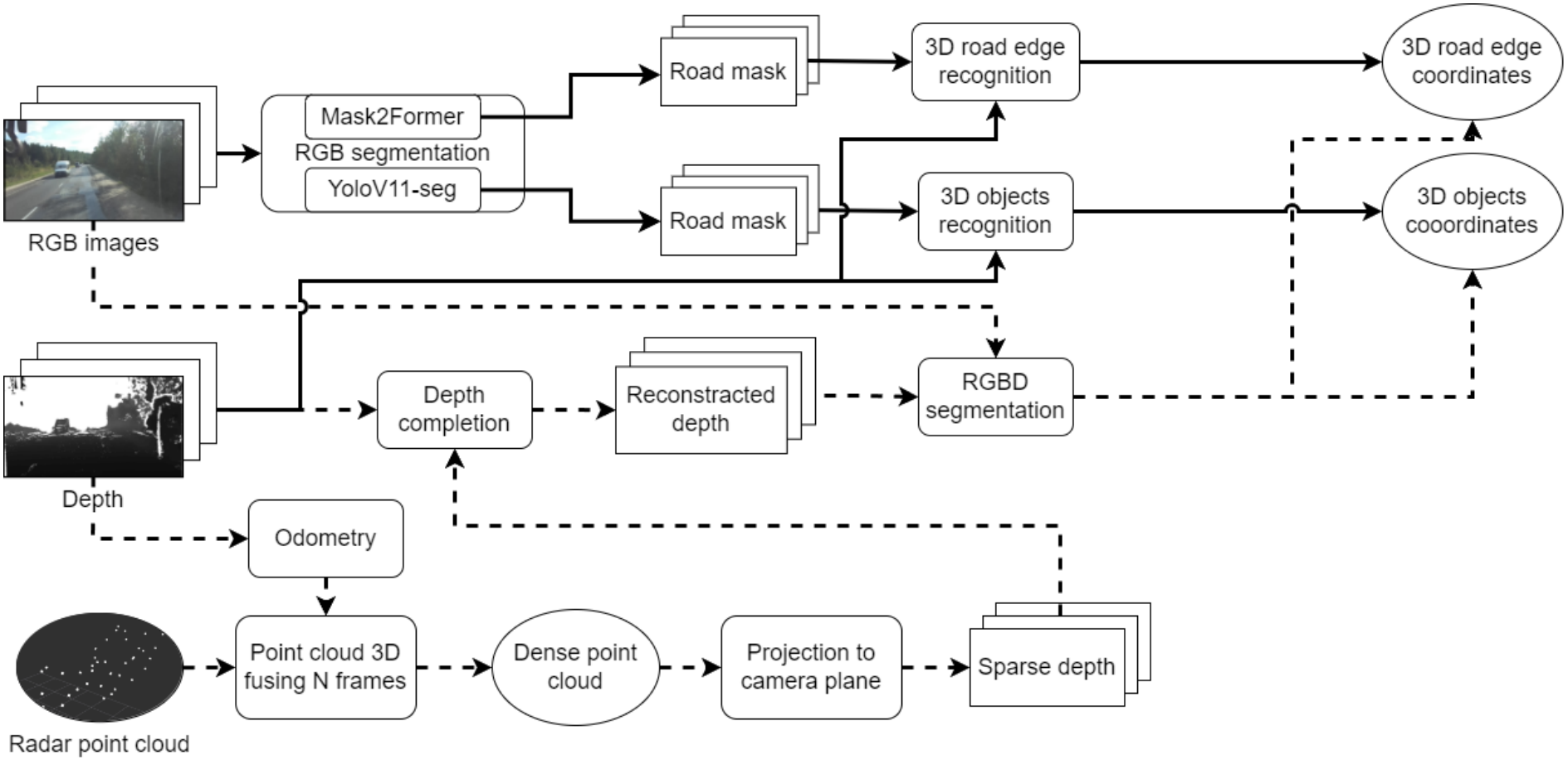
Мультимодальные датасеты

Собранный датасет

- Сенсоры: стерео-камера, радар
- Содержит данные, собранные на целевой платформе
- Включает последовательности, собранные на бездорожье
- около 4 часов неразмеченных данных



Предлагаемый метод



Модель сегментации

Базовый метод

InconSeg

+ NPO - датасет подходящего домена
+ сегментирует положительные и отрицательные препятствия
+ использует RGB+Depth



Результат предобученной модели InconSeg

Модель сегментации

Базовый метод



Результат предобученной модели InconSeg на собранных в Чехове данных

Модель сегментации

RGB сегментация

В качестве модели сегментации был выбран Mask2Former[2], так как он показал лучший результат при решении похожей задачи в городской среде

Таблица: Metrics for semantic segmentation models

model	mIoU	IoU breakroad	IoU road
DeepLabV3+	53.7	33.15	58.17
SegFormer	67.32	42.49	80.6
Mask2Former	**75.66**	**53.68**	**83.58**
OneFormer	74.10	52.94	76.37
SAN	62.37	41.21	56.5

[2] Bowen Cheng и др. «Masked-attention Mask Transformer for Universal Image Segmentation». В: 2022.

Модель сегментации

RGB сегментация

Mask2Former

сегментация дороги и препятствий

YOLO11-seg+Mask2Former

сегментация препятствий



Заключение

- Собран датасет в подходящем домене со всеми необходимыми сенсорами
- Полученную путем RGB-сегментации разметку на собранном датасете можно использовать для обучения методов RGBD-сегментации

Спасибо за внимание!

matykina.ov@phystech.edu