**«Санкт-Петербургский государственный электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)»**

**(СПбГЭТУ «ЛЭТИ»)**

ОТЧЁТ ПО УЧЕБНОЙ ПРАКТИКЕ (НИР)

Тема: **РАЗРАБОТКА ИНСТРУМЕНТА СЕГМЕНТАЦИИ 3D-МОДЕЛЕЙ ПО СНИМКАМ ДРОНОВ**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Студент |  |  |  | Курлин Н.Н. |
|  |  | *подпись* |  |  |
| Руководитель | к.т.н. |  |  | Заславский М.М. |
|  |  | *подпись* |  |  |
|  |

Санкт-Петербург

2022

СОДЕРЖАНИЕ

[ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ 3](#_Toc122145873)

[ВВЕДЕНИЕ 4](#_Toc122145874)

[1 Укрупнённое описание постановки задачи 5](#_Toc122145875)

[1.1 Актуальность 5](#_Toc122145876)

[1.2 Проблема 5](#_Toc122145877)

[1.3 Объект исследования 5](#_Toc122145878)

[1.4 Предмет исследования 5](#_Toc122145879)

[1.5 Цель работы 5](#_Toc122145880)

[1.6 Задачи 6](#_Toc122145881)

[2 ОБЗОР ЛИТЕРАТУРЫ 6](#_Toc122145882)

[2.1 Классификация методов 6](#_Toc122145883)

[2.2 Описание принципа отбора аналогов 7](#_Toc122145884)

[2.3 Сравнительный анализ методов 7](#_Toc122145885)

[3 План работы на весенний семестр 10](#_Toc122145886)

[ЗАКЛЮЧЕНИЕ 11](#_Toc122145887)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 12](#_Toc122145888)

#### ОПРЕДЕЛЕНИЯ, ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

Семантическая сегментация – это задача объединения в кластеры частей изображения, принадлежащих к одному и тому же классу объектов [1];

2D-CNN – 2D convolutional neural network [2];

Воксел – аналог двумерного пикселя в трёхмерном пространстве [3];

OA – overall accuracy (общая точность);

#### ВВЕДЕНИЕ

Уменьшение аварийности на дорогах является одной из важных задач для человеческого благополучия. Помимо принятых мер по введению ПДД, необходимо также следить за их соблюдением.

Частой и опасной проблемой является неправильная парковка транспортных средств. Предлагается использовать дронов для получения снимков улиц города и их дальнейшего объединения в 3D модель, на которой описанной в этой работе моделью можно будет обнаружить автомобили.

# Укрупнённое описание постановки задачи

## Актуальность

По статистике, опубликованной 25 июня 2021 года, 58% опрошенных ответили, что в их семье есть один автомобиль, ещё у 15% - два и более [4]. Автомобили прочно заняли место в жизни человека и принесли в неё много комфорта, однако вместе с тем и много проблем. Одной из них является проблема неправильной парковки транспортных средств. Например, в период с января по октябрь 2020 года в Москве было эвакуировано более 27 тысяч неправильно припаркованных автомобилей, оставленных на тротуарах, пешеходных переходах и остановках [5]. Неправильно припаркованные автомобили закрывают обзор, мешают пешеходам, препятствуют проезду автомобилей специальных служб, становятся причинами ДТП [6,7]. На данный момент меры по этой проблеме приняты, в основном, санкционные (штраф), необходимо также принять меры по обнаружению неправильно припаркованных автомобилей.

## Проблема

Уменьшение аварийности на дорогах

## Объект исследования

Объектом исследования являются трёхмерные модели пространства

## Предмет исследования

Точность семантической сегментации 3D моделей

## Цель работы

Разработка инструмента, обнаруживающего неправильно припаркованные автомобили по снимкам с дронов.

## Задачи

* Обзор предметной области
* Формулировка требований к решению
* Описание решения
* Оценка точности

# ОБЗОР ЛИТЕРАТУРЫ

## Классификация методов

Все методы семантической сегментации 3D модели делятся по типу входных данных [8] на методы, работающие с многоракурсными изображениями или делающие 2D срезы (multiview-based) [9], оперирующие облаком точек (point-based) [10] или вокселями (voxel-based) [11].

**Multiview-based**

Методы, принимающие на вход многоракурсные изображения или работающие с 2D срезами, обрабатывают их при помощи 2D свёрточной нейронной сети (2D-CNN) [2] и проецируют результаты сегментации в трёхмерное пространство. Методы этого вида имеют существенный недостаток: поскольку 2D изображения являются лишь приближением к 3D сценам, проявляются многочисленные ограничения и потери в геометрических структурах [12].

**Voxel-based**

Методы, использующие воксельное представление 3D модели, могут дополнительно обработать вокселизированные данные с помощью 3D-CNN. По сравнению с облаком точек, воксельная структура имеет низкое разрешение и присутствуют потери в представлении данных. Кроме того, в воксельных структурах хранятся не только занятые места, но и свободные или неизвестные места, что может привести к высоким требованиям к вычислениям и памяти [12].

**Point-based**

Методы принимают необработанные данные облака точек и генерируют окончательный прогноз [12]. Однако облака точек неупорядочены и неструктурированы, что делает невозможным прямое применение стандартных CNN [14].

## Описание принципа отбора аналогов

Поскольку сегментация будет осуществляться по снимкам с дронов, то необходимо, чтобы метод принимал на вход изображения или их производные. Важно уесть, что аналоги должны быть способны проводить именно семантическую сегментацию [13], то есть должны быть способны не только разделить объекты на снимках, но и определить для каждого из них класс (дерево, машина, дом). Метод необходимо выбрать такой, чтоб он был способен распознать объекты, как при разреженной плотности объектов, так и при высокой плотности.

Перечисленным требованиям удовлетворяют следующие методы: LDGCNN [15] SnapNet [16], PointNet++ [17].

## Сравнительный анализ методов

**Методы**

*LDGCNN* – оптимизация методов DGCNN [18] и PointNet [19]. Принимает на вход облако точек. Сначала по облаку точек строится ориентированный граф, затем из построенного графа извлекаются признаки, по которым осуществляется классификация и сегментация исходного облака точек. Обучение делится на два этапа: обучение модели, которая извлекает признаки, и обучение классификатора при фиксированной модели для извлечения признаков.

Модель лучше сходится в процессе обучения и имеет меньшее число параметров по сравнению с DGCNN.

Модель обучена на наборах данных: ModelNet40 [20], ShapeNet [21].

OA = 92,9 % [15].

*SNAPNet –* multiview-based модель, основанная на генерации по облаку точек двух типов изображений с различными положениями камеры: rgb и изображение глубины (рисунок 1). Семантическая сегментация определяет класс каждой пары соответствующих пикселей из двух входных изображений. Далее семантически сегментированные изображения проецируются в 3D.

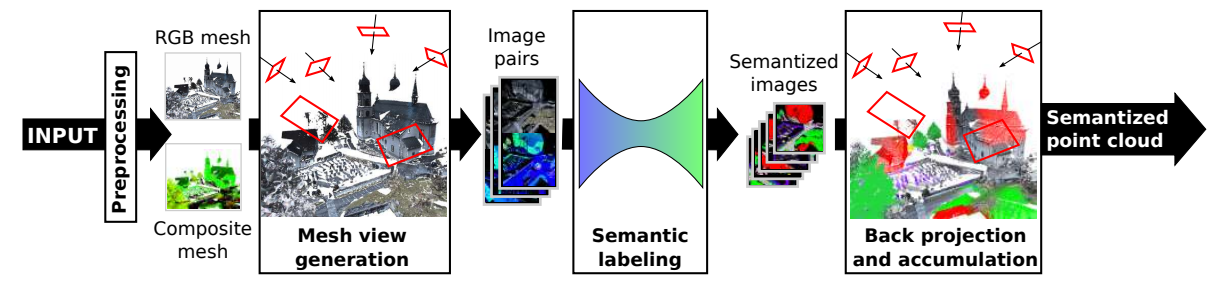


Рис. 1 – Схема работы модели SNAPNet

Модель обучена на наборе данных: Semantic 3D dataset [22].

OA = 91,0 % [14].

*PointNet++ –* расширение PointNet [23] с добавлением иерархической структуры. Сеть строит иерархическую группировку точек и постепенно абстрагирует все более и более крупные локальные области в иерархии. Иерархическая структура состоит из ряда заданных уровней абстракции. На каждом уровне набор точек обрабатывается и абстрагируется для создания нового набора с меньшим количеством элементов. Заданный уровень абстракции состоит из трех ключевых слоев: слоя выборки, слоя группировки и слоя PointNet. Слой Sampling выбирает набор точек из входных точек, которые определяют центроиды локальных областей. Затем группирующий слой строит наборы локальных областей, находя «соседние» точки вокруг центроидов. Слой PointNet использует мини-PointNet для кодирования шаблонов локальной области в векторы признаков.

Модель обучена на наборе данных: Semantic 3D dataset [22].

OA = 85,7 % [14].

**Критерии**

*Датасеты для обучения* – является важным критерием, поскольку, чем больше данных представлено для обучения модели, тем точнее она распознаёт объекты. Кроме того, большее число датасетов расширит набор классов, которые модель может распознать.

*OA (overall accuracy)* – определяет точность распознавания объектов моделью. Измеряется в процентах, определяется, как отношение числа правильно классифицированных точек к числу всех точек в облаке.

*Число классов –* какмного классов способна сегментировать модель. Напрямую зависит от датасетов для обучения. Важный критерий, т.к. может возникнуть необходимость для сегментирования новых объектов; если модель распознаёт большое число классов, то с большей вероятностью она уже умеет распознавать необходимый класс.

Сравнение методов представлено в таблице 1.

Таблица 1 – Сравнение методов по критериям.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Критерий | | |
| Метод | Датасеты для обучения | OA | Число классов |
| *LDGCNN* | ModelNet40, ShapeNet | 92,9 % | Более 900 |
| *SNAPNet* | Semantic 3D dataset | 91,0 % | 8 |
| *PointNet++* | Semantic 3D dataset | 85,7 % | 8 |

**Выводы**

Среди всех моделей наибольшей точностью по критерию OA обладает модель, основанная на использовании свёрток на графах LDGCNN. По критерию датасетов модель LDGCNN также превосходит аналоги, поскольку датасеты ModelNet40 и ShapeNet содержат более 600 и 300 классов объектов соответственно и более 150 000 3D моделей совместно, в то время как Semantic 3D dataset хоть и содержит крупные модели облаков точек (более 20 млн точек), но определяет всего 8 классов (искусственная местность, природная местность, низкая растительность, здания, машины и др.).

# План работы на весенний семестр

В весеннем семестре необходимо выбрать датасет с фотографиями улиц с машинами с воздуха и получить из них 3D модель. Запустить семантическую сегментацию с помощью каждой из моделей, сравнить результаты. Если будут обнаружены новые критерии для сравнения либо новые модели, дополнить раздел «Сравнительный анализ методов».

#### ЗАКЛЮЧЕНИЕ

В ходе работы даны ответы на ключевые вопросы исследования, проведён поиск аналогов, их анализ и сравнение.

#### СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Semantic Segmentation | Papers With Code [Электронный ресурс]. URL: <https://paperswithcode.com/task/semantic-segmentation> (дата обращения: 17.12.2022).
2. Gu J. et al. Recent advances in convolutional neural networks //Pattern recognition. – 2018. – Т. 77. – С. 354-377.
3. Voxel – Wikipedia [Электронный ресурс]. URL: <https://en.wikipedia.org/wiki/Voxel> (дата обращения: 16.12.2022).
4. Свободных мест нет. Россияне — о проблемах с парковками в стране | Анкетолог [Электронный ресурс]. URL: <https://iom.anketolog.ru/2021/06/25/parkovki> (дата обращения: 08.12.2022).
5. В Москве раскрыли статистику эвакуации неправильно припаркованных машин :: Autonews [Электронный ресурс]. URL: <https://www.autonews.ru/news/5f7578219a7947553e62f220> (дата обращения: 08.12.2022).
6. Заблокированная «скорая» с ребенком стала федеральным скандалом (видео) [Электронный ресурс]. URL: <https://chelny-biz.ru/news/431338/> (дата обращения: 08.12.2022).
7. Краснодар | В Краснодаре неправильно припаркованный автомобиль спровоцировал ДТП ВИДЕО - БезФормата [Электронный ресурс]. URL: <https://krasnodar.bezformata.com/listnews/avtomobil-sprovotciroval-dtp/96719818/> (дата обращения: 08.12.2022).
8. Xie Y., Tian J., Zhu X. X. Linking points with labels in 3D: A review of point cloud semantic segmentation //IEEE Geoscience and Remote Sensing Magazine. – 2020. – Т. 8. – №. 4. – С. 38-59.
9. Su H. et al. Multi-view convolutional neural networks for 3d shape recognition //Proceedings of the IEEE international conference on computer vision. – 2015. – С. 945-953.
10. Engelmann F. et al. Exploring spatial context for 3D semantic segmentation of point clouds //Proceedings of the IEEE international conference on computer vision workshops. – 2017. – С. 716-724.
11. Maturana D., Scherer S. Voxnet: A 3d convolutional neural network for real-time object recognition //2015 IEEE/RSJ international conference on intelligent robots and systems (IROS). – IEEE, 2015. – С. 922-928.
12. Xie Y., Tian J., Zhu X. X. Linking points with labels in 3D: A review of point cloud semantic segmentation //IEEE Geoscience and Remote Sensing Magazine. – 2020. – Т. 8. – №. 4. – С. 38-59.
13. Garcia-Garcia A. et al. A review on deep learning techniques applied to semantic segmentation //arXiv preprint arXiv:1704.06857. – 2017.
14. Guo Y. et al. Deep learning for 3d point clouds: A survey //IEEE transactions on pattern analysis and machine intelligence. – 2020. – Т. 43. – №. 12. – С. 4338-4364.
15. Zhang, Kuangen and Hao, Ming and Wang, Jing and de Silva, Clarence W. and Fu, Chenglong. Linked dynamic graph cnn: learning on point cloud via linking hierarchical features // arXiv:1904.10014. – 2019.
16. A. Boulch, J. Guerry, B. Le Saux, and N. Audebert. Snapnet: 3d point cloud semantic labeling with 2d deep segmentation networks //Computers & Graphics. – vol. 71, pp. 189–198. – 2018.
17. C. R. Qi, L. Yi, H. Su, and L. J. Guibas. PointNet++: Deep hierarchical feature learning on point sets in a metric space //NeurIPS. – 2017.
18. Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon. Dynamic Graph CNN for Learning on Point Clouds //arXiv:1801.07829 [cs]. – Jan. 2018.
19. C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep Learning on Point Sets for 3d Classification and Segmentation //IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Jul. 2017. – pp. 77–85.
20. Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d ShapeNets: A deep representation for volumetric shapes //IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – Jun. 2015. – pp. 1912–1920.
21. L. Yi, V. G. Kim, D. Ceylan, I.-C. Shen, M. Yan, H. Su, C. Lu, Q. Huang, A. Sheffer, and L. Guibas. A Scalable Active Framework for Region Annotation in 3d Shape Collections //ACM Trans. – 2016. – Graph., vol. 35, no. 6, pp. 210:1–210:12.
22. HACKEL T., SAVINOV N., LADICKY L., WEGNER J.-D., SCHINDLER K., POLLEFEYS M. Large-scale point cloud classification benchmark //CVPR/ Large Scale 3D Data Workshop. – 2016.
23. C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation //arXiv preprint arXiv:1612.00593. – 2016.