

## НЕЙРОСЕТЕВЫЕ ТЕХНОЛОГИИ РАСПОЗНАВАНИЯ ТРЕХМЕРНЫХ ОБЪЕКТОВ

А.М. Бабаев, студент

Институт сферы обслуживания и предпринимательства (филиал) ДГТУ  
(Россия, г. Шахты)

DOI: 10.24411/2500-1000-2019-11889

**Аннотация.** Статья посвящена рассмотрению методов распознавания трехмерных объектов. Приведены примеры наиболее известных архитектур сверточных нейронных сетей, применяемых при анализе трехмерных сцен. Перечислены основные форматы представления трехмерных данных. Приведены результаты тестирования описанных в статье методов на наборе данных, содержащем трехмерные облака точек.

**Ключевые слова:** нейронные сети, трехмерное облако точек, воксельная сетка, распознавание объектов, сегментация объектов.

Распознавание объектов является важной частью современных интеллектуальных систем. Наиболее эффективным решением этой задачи на данный момент являются сверточные нейронные сети (далее СНС), которые стали одной из причин прорыва в области компьютерного зрения [1]. Наиболее известными являются архитектуры нейронных сетей, которые работают с двумерными изображениями. Однако различные модели применяются и при работе с трехмерными данными.

С повышением уровня доступности стереокамер и LiDAR-датчиков исследования в области распознавания трехмерных объектов привлекают все больше исследователей. Уже сейчас существует ряд технологий, позволяющих решать задачи классификации и реконструкции трехмерных объектов, сегментации и семантического анализа трехмерных сцен. В данной работе будут рассмотрены наиболее популярные нейросетевые методы распознавания трехмерных объектов.

Трехмерная сцена при построении с помощью любого сканирующего устройства представлена в виде трехмерного облака точек [2]. При использовании облаков точек в распознавании объектов возникает ряд проблем:

– неупорядоченность – облако точек представляет собой набор точек без определенного порядка;

– между точками имеется корреляция в виде расстояния между точками, которую

необходимо каким-то образом сообщать сети;

– отсутствующие данные и шум.

Для решения этих проблем могут быть применены следующие подходы:

1. Сортировка точек. Это вычислительно неэффективный метод.

2. Использование симметричной функции для агрегирования информации. То есть такой функции, значение которой не меняется в зависимости от порядка элементов.

Этот подход использован в архитектуре PointNet. В ней была применена операция субдискретизации (pooling), а именно max pooling, который выступает для СНС базовым слоем и является симметричной функцией.

Авторы архитектуры PointNet предложили передать информацию о точках (x, y, z) напрямую в глубокую нейронную сеть [3]. Однако мощности персональных компьютеров недостаточно для анализа миллионов точек. Поэтому было предложено сначала вычислять по двумерному изображению примерное расположение объекта, а потом вырезать участок облака точек и анализировать только его. Сеть классификации принимает в качестве входных данных n точек, применяет преобразования входных данных и признаков, а затем агрегирует признаки точек с помощью max pooling. Глобальный вектор признаков идет на вход полносвязному слою для определения класса всего облака

точек. Результатом является классификационная оценка для каждого класса.

В отличие от 2D-изображений, при работе с трехмерными сценами до сих пор не решен вопрос о том, какой формат ввода данных лучше использовать непосредственно в сверточных нейронных сетях. Наряду с облаком точек используется воксельная сетка (Voxel Grid) – это наиболее интуитивный метод, который встраивает трехмерные объекты в сетку, чтобы она

выглядела как пиксельное изображение. Ее можно получить из облака точек (рисунок). Для этого необходимо создать массив  $32 \times 32 \times 32$ , который будет заполнен нулями. Затем масштабируем облако точек и вычисляем, какие точки находятся внутри каждого маленького вокселя. После этого каждому вокселю назначается цвет как среднее арифметическое точек внутри него, либо берется самый часто встречающийся цвет.



Рисунок. Преобразование облака точек в воксельную сетку

В качестве нейронной сети, работающей с воксельной сеткой, рассмотрим трехмерную сверточную нейронную сеть VoxNet для распознавания объектов в реальном времени. Она была предложена авторами в качестве технологии распознавания автомобилей и пешеходов [4]. Архитектура состоит из 3-х основных частей – сети обучения функций (Feature Learning Network), сверточных средних слоев (Convolutional Middle Network) и сети предложения регионов (Region Proposal Network).

Сеть обучения признаков сначала выполняет вокселизацию пространства. Поскольку 90% вокселей являются пустыми, этот слой учитывает разреженную тензорную обработку посредством обработки только непустых вокселей.

После получения воксельной сетки к ней необходимо применить трехмерные свертки. Их основное отличие от двумерной операции свертки состоит в том, что в качестве ядра весов используется не квадратная матрица, а трехмерная матрица-куб.

RPN, является последней стадией обработки и состоит из 3 подблоков двумерных

сверточных слоев, за которыми следуют 3 блока двумерных свертки.

Использование воксельной сетки является наиболее простым и интуитивно понятным методом распознавания трехмерных объектов. Но при использовании воксельной сетки возникает ряд проблем. Во-первых, при масштабировании облака точек может возникнуть ситуация, когда одному вокселю будут соответствовать точки целого объекта. В таком случае задача распознавания образов не может быть адекватно решена. Во-вторых, нелегко определить цвет RGB для каждого вокселя из-за операций масштабирования облака точек.

Таким образом, для набора данных с одинаковым количеством точек и схожим масштабом сканирования VoxNet является простым и быстрым решением. Однако при наличии сложного набора данных воксельная сетка не является лучшим выбором.

Для тестирования архитектур PointNet и VoxelNet был использован набор данных KITTI [5]. Он состоит из 7481 обучающих и 7518 тестовых сцен, включающих в общей сложности 80 256 помеченных объек-

тов. Облака точек разделены на три группы в зависимости от сложности ситуации распознавания:

1. Легкий случай: мин. высота ограничительной рамки: 40 пикселей, макс. уровень загромождения: полностью видимый, макс. усечение рамки: 15%.

2. Умеренный случай: мин. высота ограничительной рамки: 25 пикселей, макс.

уровень загромождения: частично закрыт, макс. усечение рамки: 30%.

3. Трудный: мин. высота ограничительной рамки: 25 пикселей, макс. Уровень загромождения: трудно увидеть, макс. усечение рамки: 50%.

Результаты тестирования для класса Cars приведены в таблице.

Таблица. Результаты тестирования

Технология	FPS	Точность, %		
		Легкий	Умеренный	Трудный
VoxelNet	4,3	89,35	79,26	73,39
PointNet	5,9	88,7	84	75,33

По результатам тестирования можно сделать следующие выводы. В простых случаях наиболее эффективно применять

архитектуру VoxelNet. А наиболее точным решением при работе в трудных условиях стала сеть PointNet.

#### Библиографический список

1. NVIDIA and IBM Cloud Support ILSVRC / Официальный сайт компании NVIDIA. – [Электронный ресурс]. – Режим доступа: <https://devblogs.nvidia.com/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/> (дата обращения: 20.12.2019).

2. *An introduction towards 3D Computer Vision* / Medium. – [Электронный ресурс]. – Режим доступа: [https://medium.com/@jjianshi\\_94445/an-introduction-towards-3d-computer-vision-71be8ce11956](https://medium.com/@jjianshi_94445/an-introduction-towards-3d-computer-vision-71be8ce11956) (дата обращения: 20.12.2019).

3. Charles, R.Q. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation / R.Q. Charles, S. Hao, M. Kaichun, J.G. Leonidas // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 652-660.

4. Maturana, D. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition / D. Maturana, S. Scherer // International Conference on Intelligent Robots and Systems (IROS). – 2015. – P. 922-928.

5. *3D Object Detection Evaluation 2017* // [Официальная страница соревнования] / The KITTI Vision Benchmark Suite. – [Электронный ресурс]. – Режим доступа: [http://www.cvlibs.net/datasets/kitti/eval\\_object.php?obj\\_benchmark=3d](http://www.cvlibs.net/datasets/kitti/eval_object.php?obj_benchmark=3d) (дата обращения: 22.12.2019).

## NEURAL NETWORK TECHNOLOGIES OF RECOGNITION OF THREE-DIMENSIONAL OBJECTS

A.M. Babaev, Student

Institute of Service and Entrepreneurship (branch) of DSTU  
(Russia, Shakhty)

**Abstract.** The article is devoted to the consideration of three-dimensional object recognition methods. Examples of the most famous convolutional neural network architectures used in the analysis of three-dimensional scenes are given. The main formats for representing three-dimensional data are listed. The results of testing the methods described in the article on a data set containing three-dimensional point clouds are presented.

**Keywords:** neural networks, three-dimensional point cloud, voxel network, object recognition, segmentation of objects.