

## СЕГМЕНТАЦИЯ ОБЛАКА ТОЧЕК С НЕИЗВЕСТНЫМИ ОБЪЕКТАМИ С ПОМОЩЬЮ МЕТОДА VCCS И ДИНАМИЧЕСКОЙ ГРАФОВОЙ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ

Воронков А.Д.

*МИРЭА - Российский технологический университет, 119454 г. Москва, проспект Вернадского, дом 78  
e-mail: a.voronkov.rtu@yandex.ru.*

---

В статье представлен метод сегментации облака точек сцены, состоящей из неизвестных объектов, на основе использования метода сегментации Voxel Cloud Connectivity Segmentation (VCCS) и двухэтапной обработки векторов признаков с применением нейронной сети PointNet и динамической графовой сверточной нейронной сети (Dynamic Graph Convolutional Neural Network, DGCNN). В ряде случаев практическое применение манипуляционных роботов сопряжено с захватом объектов, форма, цвет и другие признаки которых не известны заранее. В частности, примерами таких задач могут служить уборка помещения, аварийно-спасательные операции по разбору завалов, работа в складских помещениях или в почтовых отделениях. В предлагаемом подходе по набору изображений загроможденной сцены, получаемому с RGBD-камер, составляется образ сцены в виде облака точек, затем данное облако точек подвергается обработке с помощью эвристического алгоритма VCCS и методов машинного обучения. Результатом работы подхода является сегментированное облако точек, для каждой точки которого имеется метка, определяющая ее принадлежность к отдельному объекту сцены. Новизна подхода заключается в совмещении эвристического алгоритма VCCS и новой архитектуры нейронной сети, представляющей собой комбинацию модифицированных сетей PointNet и DGCNN. Проведенные экспериментальные исследования подтверждают работоспособность предлагаемого решения.

---

Ключевые слова: машинное обучение, сегментация облака точек, графовая нейронная сеть, сегментация неизвестных объектов, супервоксели, VCCS, PointNet, DGCNN.

## SEGMENTATION OF A POINT CLOUD WITH UNKNOWN OBJECTS USING THE VCCS METHOD AND A DYNAMIC GRAPH CONVOLUTIONAL NEURAL NETWORK

Voronkov A.D.

*MIREA — Russian Technological University, 119454, Russia, Moscow, Vernadsky Avenue, 78, e-mail:  
a.voronkov.rtu@yandex.ru.*

---

The article presents a method for segmenting a point cloud of a scene consisting of unknown objects based on the use of the Voxel Cloud Connectivity Segmentation (VCCS) method and two-stage feature vector processing using the PointNet neural network and Dynamic Graph Convolutional Neural Network (DGCNN). In some cases, the practical application of manipulative robots involves grasping objects whose shape, color and other features are not known in advance. In particular, examples of such tasks can be cleaning of premises, emergency rescue operations to remove blockage, work in warehouses or in post offices. In the proposed approach, an image of the scene in the form of a point cloud is compiled from a set of images of a cluttered scene obtained from RGBD cameras, then this point cloud is processed using the VCCS heuristic algorithm and machine learning methods. The result of the approach is a segmented point cloud, for each point of which there is a label that determines its belonging to a separate object in the scene. The novelty of the approach lies in the combination of the VCCS heuristic algorithm and the new neural network architecture, which is a combination of modified PointNet and DGCNN networks. The conducted experimental studies confirm the operability of the proposed solution.

---

Keywords: machine learning, point cloud segmentation, graph neural network, unknown objects segmentation, supervoxels, VCCS, PointNet, DGCNN.

### Введение

В современном мире манипуляционная робототехника играет важную роль: роботы участвуют в сборке техники, проведении аварийно-спасательных операций, сборе урожая. Вместе с тем в настоящее время

наблюдается рост качества работы и снижение стоимости RGBD-камер, которые позволяют получать изображения, содержащие 4-канальный массив из RGB-изображения и карты глубины. Таким образом, круг задач, которые могут быть решены с помощью манипуляционных роботов, постепенно расширяется.

На сегодняшний день хорошо изучены и испытаны на большом количестве прикладных задач алгоритмы планирования целенаправленных движений манипуляционных роботов в средах с препятствиями. Целый ряд исследований посвящен задаче выбора наилучшего положения захватного устройства, максимизирующего вероятность успешного захвата и перемещения объекта в рабочей области манипулятора. Вместе с тем, поскольку современные задачи автоматизации и робототехники во многом опираются на системы восприятия, основанные на применении RGBD-камер, были созданы широко известные открытые библиотеки для работы с 3D-данными, машинным обучением и компьютерным зрением, такие как Point Cloud Library (PCL), Open3D, Scikit Learn, PyTorch и др. Их применение позволяет разрабатывать эффективные алгоритмы обработки данных с камеры робота.

В настоящее время одной из ключевых задач в сфере восприятия манипуляционным роботом рабочей среды является задача выделения заранее неизвестных объектов сцены. Поскольку сегментация облака точек – фундаментальный шаг для анализа сцены, разработка методов сегментации объектов априорно неизвестной формы на основе анализа признаков их видимой поверхности позволит манипуляционным роботам с большей эффективностью функционировать в неопределенных заранее средах, таких как улица и новое помещение.

### **Обзор существующих разработок предметной области**

Автоматическая сегментация является частью задачи по восприятию, оценке и интерпретации сцены (англ. scene understanding). В отличие от сегментации 2D-изображений, где сегодня достигнуты большие успехи за счет применения сверточных нейронных сетей, сегментация 3D-облаков точек находится на стадии активных исследований. Особую сложность в решении данной задачи представляет загроможденность сцены, то есть случай, когда объекты сцены расположены близко друг к другу или друг на друге.

Прикладная задача сегментации сцены часто предполагает наличие базы данных с априорно известным представлением объектов рабочей области. Так, в [1] использовалась математическая модель объектов в виде упорядоченного набора данных. Однако в условиях неопределенности робот может столкнуться с большим числом априорно неизвестных объектов, для которых заранее не известна их форма, цвет, кривизна и другие внешние признаки. Поэтому в научной литературе представлено несколько вариантов решения задачи сегментации сцены с объектами произвольной формы. Данные решения характеризуются совокупностью критериев, такими как вычислительная эффективность и точность.

Задача сегментации объектов произвольной формы особенно актуальна в задаче автоматического захвата. Так, существующие методы захвата неизвестных объектов часто предполагают работу на плоской поверхности [2]. Используя итеративный метод оценки параметров математической модели по случайной выборке из данных RANdom SAmple Consensus (RANSAC), в данных подходах производится поиск плоскости, на которых расположены объекты. Затем точки плоскости удаляются, а оставшиеся подмножества точек подвергаются алгоритму кластеризации, который выдает кластеры точек, принадлежащие отдельным объектам. Данный подход распространен благодаря своей простоте и эффективности, однако он не может быть применен, чтобы сегментировать загроможденную сцену, где объекты расположены близко друг к другу или где нет видимой поверхности стола.

Алгоритмы компьютерного зрения, широко используемые в прикладных задачах манипуляционной и мобильной робототехники, также могут быть применены в задаче сегментации сцены с произвольными объектами. В [3] для выделения отдельных объектов, расположенных на поверхности стола, плоскость выделялась на основе анализа ее цвета. Оставшиеся объекты сегментировались при помощи алгоритма водораздела, который, на основе информации об интенсивности пикселей, объединял однородные области изображения. Несмотря на то, что данный метод может приводить к чрезмерной сегментации, то есть к разделению одного и того же объекта на несколько сегментов, он применим в задаче автоматизированного захвата неизвестных объектов.

В [4] для сегментации сложной визуальной сцены на поверхности стола использовалось разбиение исходного облака точек на супервоксели при помощи алгоритма растущих областей Voxel Cloud Connectivity Segmentation (VCCS) [5]. Полученные супервоксели при помощи иерархической кластеризации последовательно объединялись в сегменты объектов сцены. Данный подход показал хорошую работу, достигнув взвешенного коэффициента Жаккара для экземпляров сцены  $IoU_w = 0.9298$ . Преимущество работы с 3D-облаком точек, преобразованным в воксельную сетку, помимо фильтрации шумов, является тот факт, что максимальное количество вокселей сцены определяется разрешением, то есть стороной вокселя. Таким образом, алгоритмы для работы с вокселями не зависят от количества исходных RGBD-изображений. Стоит также отметить, что

при использовании представления окружающей среды в виде облака точек или трехмерной сетки вокселей не возникает потребности в совмещении сегментов, выделенных на нескольких 2D-изображениях.

В [6] авторы использовали модифицированную версию нейросетевой архитектуры Mask R-CNN для сегментации неизвестных объектов на изображениях глубины. Для этого на вход нейронной сети подавалось изображение глубины, преобразованное в трехканальное черно-белое изображение, количество классов было понижено до двух: класса фона и класса объекта переднего плана. Нейронная сеть обучалась на выборке, созданной с помощью виртуального моделирования с помощью библиотеки PyBullet. В результате нейронная сеть обучалась выделять сегменты экземпляров объектов на изображении независимо от их класса. Данное исследование показывает, что обобщающие способности нейронной сети могут быть использованы для усваивания концепции экземпляра класса «объект».

В [7] исследовался метод растущих областей с применением нейронной сети. Поскольку в условиях реального мира облака точек часто содержат объекты произвольной формы и внешнего вида, целью авторов было произвести сегментацию исходного облака точек на экземпляры объектов, используя нейросетевую архитектуру LRGNet. Нейронная сеть LRGNet использовалась итерационно в процессе увеличения области, определяя точки, которые нужно добавить или удалить из текущей области, достигнув среднего коэффициента Жаккара на наборе данных S3DIS  $IoU_m=0.54$ . Недостатком метода является сравнительно большое время работы по сравнению с другими алгоритмами из-за необходимости многократно запускать нейронную сеть.

В [8] для сегментации RGBD-изображения с неизвестными объектами использовался нейросетевой подход на базе применения полностью сверточной нейронной сети с энкодер-декодерной архитектурой. Авторы использовали обучение сети с функцией потерь, которая стимулировала сеть объединять в кластеры векторы признаков каждого пикселя одного и того же объекта и отталкивать центроиды разных кластеров друг от друга. За счет использования 2-этапного алгоритма выделения признаков и их кластеризации была достигнута F1-мера 91.6 на наборе данных OCID и 87.4 на наборе данных OSD.

Для анализа 3D-облаков точек был создан ряд нейросетевых архитектур, таких как PointNet [9], PointNet++ [10] и DGCNN [11], которые показали хорошие результаты в задачах классификации и сегментации облаков точек. Нейронная сеть PointNet обрабатывает каждую точку в исходном облаке точек отдельно с помощью многослойного перцептрона (МСП). Полученные таким образом векторы признаков точек далее подвергаются операции выборки максимального значения, формируя вектор глобальных признаков облака точек. Архитектура PointNet++ имеет структуру энкодер-декодер. Для кодирования признаков на разных уровнях используются слои абстракции, а для распространения признаков в декодере используются слои распространения признаков. Нейронная сеть DGCNN представляет облако точек в виде ориентированного графа  $G = (V, E)$ , где  $V$  – узлы графа, соответствующие точкам, и  $E$  – матрица смежности графа. В [11] авторы также представили слой EdgeConv, реализующий выделение набора признаков узла графа, содержащих как глобальный, так и локальный контекст.

Подводя итог обзора существующих разработок предметной области, можно сделать вывод, что в сфере манипуляционной робототехники существует запрос на разработку методов сегментации произвольных объектов в условиях реального мира. Одними из самых новых, но всецело не изученных подходов к решению данной задачи является выделение неявных признаков точек изображения или облака точек, и их последующая кластеризация. Таким образом, проблема сегментации облака точек с априорно неизвестными объектами является актуальной.

### **Постановка задачи исследования**

Исходное неорганизованное облако точек описывается конечным множеством  $P = \{p_1, p_2, \dots, p_n\}$  с мощностью  $n$ , каждый элемент  $p_i$  которого является точкой в пространстве признаков. В пространстве  $R^9$  признаков каждая точка описывается 9 координатами  $(X, Y, Z, C_R, C_G, C_B, N_x, N_y, N_z)$ , где  $(X, Y, Z)$  – координаты точки  $p_i$  в евклидовом пространстве  $R^3$ ,  $(C_R, C_G, C_B)$  – координаты  $p_i$  в пространстве цветов RGB,  $(N_x, N_y, N_z)$  – единичный вектор нормали к поверхности в точке  $p_i$ .

Алгоритм сегментации VCCS производит преобразование исходного облака точек  $P$  в множество вокселей  $V$ , каждый элемент в котором является точкой в пространстве признаков, где признаки для вокселя получаются путем взятия среднего арифметического для координат в евклидовом и цветовом пространстве и путем нормализации суммы векторов единичных нормалей для всех точек из множества  $P$ , находящихся внутри вокселя  $v_i$ . В процессе работы алгоритма VCCS точки  $v \in V$  преобразуются в  $m$  кластеров. Таким образом, формируется множество супервокселей  $S = \{s_1, s_2, \dots, s_m\}$ . В результате подвыборки определяется множество  $S' \subseteq S$ , состоящее из 160 супервокселей.

Результатом работы метода является определение нового признака  $L$  для каждого супервокселя  $s \in S'$  в виде метки экземпляра, указывающей, к какому экземпляру объекта принадлежит супервоксель. Таким образом,

задача разрабатываемого алгоритмического комплекса – извлечение признаков с локальным и глобальным контекстом из набора супервокселей и кластеризация супервокселей в пространстве признаков с целью сегментации объектов.

### Описание и реализация метода

Общая схема разработанного метода представлена на рис. 1. Как видно из рисунка, предлагаемый подход состоит из 4 основных шагов: сегментации исходного облака точек с помощью метода VCCS, формирования набора из 160 супервокселей, использования нейронных сетей MiniPointNet и MiniDGCNN для извлечения признаков и кластеризации полученных векторов признаков с помощью алгоритма DBSCAN.

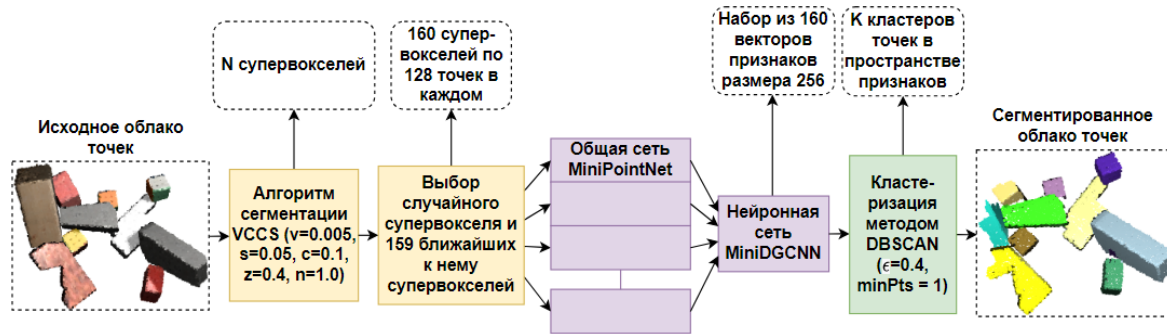


Рисунок 1. Общая схема разработанного комплекса

Методы обработки облаков точек, основанные на применении алгоритма VCCS [5], как правило, используют пересегментацию исходного облака точек для снижения размерности и фильтрации исходных данных [12]. В основе алгоритма VCCS лежит подход к выделению однородных участков облака точек путем кластеризации точек в пространстве признаков. Супервоксели являются аналогом суперпикселей, использующихся в алгоритмах обработки 2D-изображений. Данный алгоритм основан на применении графа, так как оперирует на воксельном представлении исходного облака точек, анализируя соседние воксели для каждого вокселя. Поскольку исходное облако точек может быть зашумлено и иметь неодинаковую плотность в разных местах, алгоритм VCCS решает эту проблему путем фильтрации облака точек благодаря преобразованию в воксельную сетку.

На основе равномерной выборки начальных центров кластеров и итерационной кластеризации вокселей вокруг центров кластеров, алгоритм VCCS выдает набор из подмножеств вокселей исходного облака точек, каждое из которых содержит воксели, объединенные общими признаками. В данном исследовании использовалась реализация алгоритма VCCS библиотеки PCL. В данной реализации симметрическая функция расстояния  $D(V_1, V_2)$  между центроидом  $V_1$  рассматриваемого супервокселя и рассматриваемого соседнего вокселя  $V_2$  является взвешенной суммой и рассчитывается в соответствии с формулами (1)-(4). В данной работе использовался размер вокселя  $v=0.005$  м, разрешение воксельной сетки для выборки центроидов кластеров  $s=0.05$  м и весовые коэффициенты  $w_e=0.4$ ,  $w_c=0.1$ ,  $w_n=1$ .

$$D = w_e D_e + w_c D_c + w_n D_n, \quad (1)$$

$$D_e = \|V_{1e} - V_{2e}\| / s, \quad (2)$$

$$D_c = \|V_{1c} - V_{2c}\| / 255, \quad (3)$$

$$D_n = 1 - |\vec{V}_{1n} \cdot \vec{V}_{2n}|, \quad (4)$$

где  $w_e, w_c, w_n$  - весовые коэффициенты,

$D_e$  – нормализованное расстояние между точками  $V_{1e}$  и  $V_{2e}$  в евклидовом пространстве  $R^3$ ,

$D_c$  – нормализованное расстояние между точками  $V_{1c}$  и  $V_{2c}$  в пространстве цветов RGB  $R^3$ ,

$D_n$  – мера расстояния между точками  $V_{1n}$  и  $V_{2n}$  в пространстве нормалей  $R^3$ .

Для формирования набора данных для подачи на вход нейронной сети был использован алгоритм выборки супервокселей, при котором случайным образом выбирался центральный супервоксель и  $k=159$  соседних супервокселей на основе оценки расстояния между центроидами в евклидовом пространстве. Для извлечения вектора признаков из каждого отдельного супервокселя выборки использовалась нейронная сеть с архитектурой MiniPointNet, основанная на исходной нейросетевой архитектуре PointNet.

Структура сети представлена на рис. 2а. Данная нейронная сеть состоит из 2 основных частей: многослойного перцептрона (МСП) и слоя максимальной подвыборки. МСП выделяет признаки каждой отдельной точки, входящей в супервоксель, в то время как слой максимальной подвыборки играет роль симметрической функции – то есть функции, инвариантной к перестановке ее аргументов. Таким образом, достигается аппроксимация функции  $f$ , определенной на множестве точек  $\{x_1, x_2, \dots, x_n\}$  [9]:

$$f(\{x_1, x_2, \dots, x_n\}) \approx g(h(x_1), h(x_2), \dots, h(x_n)) \quad (5)$$

где  $f$  – целевая симметрическая функция, ставящая в соответствие множеству из  $n$  аргументов пространства  $R^D$  вектор пространства  $R^L$ , где  $D$  – размер вектора признаков каждой точки,  $L$  – размер извлеченного вектора признаков облака точек,  $n$  – количество точек в облаке точек,

$h$  – функция, сопоставляющая в соответствие входной переменной пространства  $R^D$  вектор признаков пространства  $R^L$  (применение МСП),

$g$  – симметрическая функция, сопоставляющая  $n$  векторов признаков пространства  $R^L$  с одним вектором пространства  $R^L$  (применение максимальной подвыборки).

Таким образом, для каждого из 160 супервокселей, состоящих из  $n = 128$  точек, нейронная сеть MiniPointNet вначале выделяет 128 векторов признаков размера  $L = 128$ , а затем с помощью максимальной подвыборки выделяет один вектор признаков точек из 128 элементов, кодирующих определенные статистические свойства точек супервокселя. Преимуществом использования сети с PointNet-подобной архитектурой является ее скорость работы и инвариантность к перестановке точек внутри облака точек.

Для выделения векторов признаков для каждого супервокселя с учетом локального и глобального контекста использовалась нейронная сеть MiniDGNN. Данная нейронная сеть основана на архитектуре динамической графовой сверточной нейронной сети DGCNN, адаптирующей методы графовых сверточных нейронных сетей для применения на облаках точек.

Архитектура использованной сети MiniDGCNN приведена на рис. 2б. Как видно из рисунка, сеть состоит из последовательности EdgeConv-слоев, с последующей конкатенацией векторов признаков, извлеченных на предыдущих этапах работы сети. В начале работы сети строится граф  $G$ , состоящий из 160 вершин, где между каждой вершиной и  $k = 8$  ближайшими в пространстве признаков к ней вершинами имеется ребро. В процессе операции свертки при помощи EdgeConv-слоя задействуется многослойный перцептрон, принимающий на вход вектор признаков данной вершины  $x_i$  и вектор признаков, описывающий локальное соотношение вершины  $x_i$  и соседней вершины  $x_j$  в виде их разности [11]:

$$e_{ij} = h_{\theta}(x_i, x_j - x_i) \quad (6)$$

где  $h_{\theta}$  – нелинейная функция, параметризованная обучаемыми параметрами сети  $\theta$ ,  
 $e_{ij}$  – промежуточный вектор признаков ребра.

Далее векторы признаков ребер агрегируются путем выборки максимального по значению  $m$ -го признака из набора  $k$  векторов и формирования нового вектора признаков вершины графа  $x_i$ :

$$x_{im} = \max_{j:(i,j) \in E} (e_{ijm}) \quad (7)$$

После операции EdgeConv, граф  $G$  строится заново, используя алгоритм поиска  $K$  ближайших соседей в новом пространстве признаков. MiniDGCNN принимает на вход 160 векторов признаков размера 128, полученных нейронной сетью MiniPointNet, и выдает 160 векторов признаков размера 256. Таким образом, разработанная архитектура нейронной сети, совмещающая в себе выделение признаков для каждого супервокселя и выделение вектора признаков, содержащего локальный и глобальный контекст, формирует массив признаков, извлекаемых из входных данных, за счет своей внутренней структуры и связанных с ней весовых коэффициентов.

Для кластеризации полученных точек в пространстве признаков использовался алгоритм DBSCAN. Данный алгоритм используется для группировки точек данных с высокой плотностью, обеспечивая при этом помехоустойчивость.

Входными данными алгоритма являются множество точек в пространстве признаков, максимальное расстояние между двумя точками одного кластера  $\epsilon$  и минимальное количество точек в радиусе  $\epsilon$ , необходимое для формирования кластера,  $minPts$ . Выбор значений параметров  $\epsilon$  и  $minPts$  оказывает существенное влияние на

результатирующие кластеры. При проведении экспериментов было обнаружено, что наилучшие результаты метод показывает при параметрах  $\varepsilon = 0.4$  и  $minPts = 1$ .

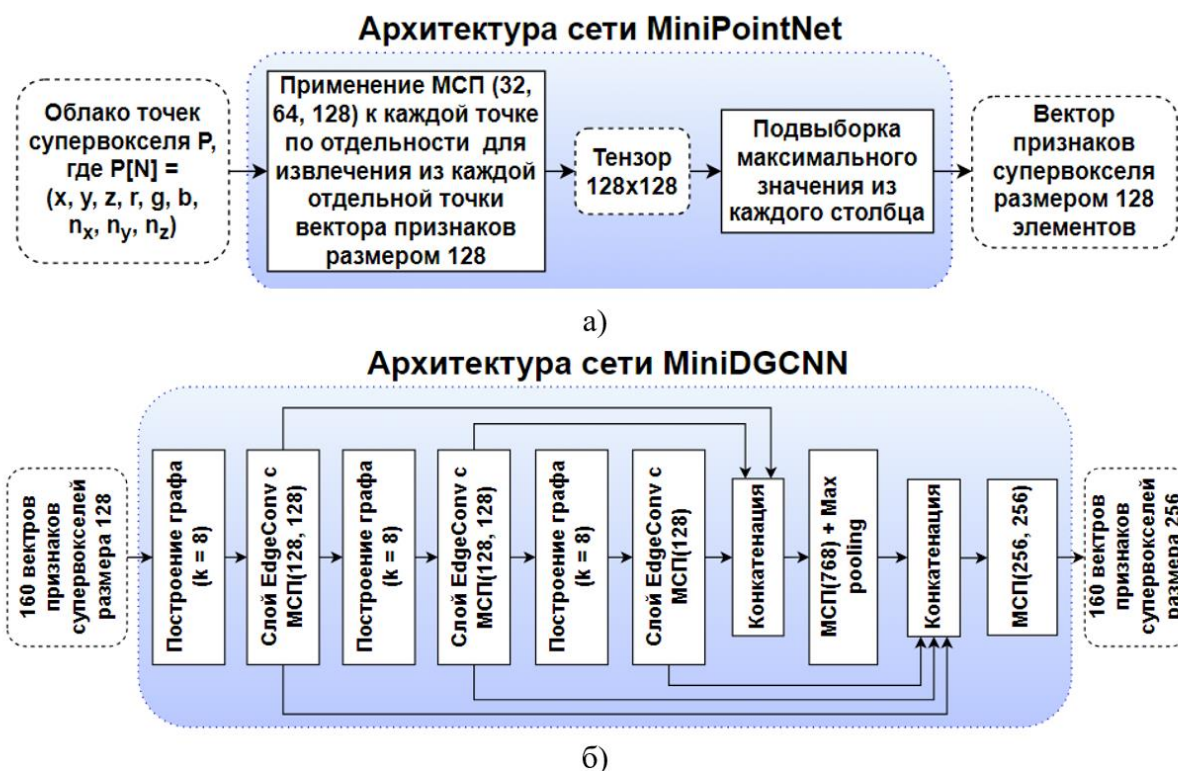


Рисунок 2. Используемые архитектуры нейронных сетей:  
а - архитектура сети MiniPointNet; б - архитектура сети MiniDGCNN

Преимуществом DBSCAN является его скорость работы, способность выделять кластеры произвольной формы и автоматическое определение количества кластеров, поскольку количество объектов сцены заранее не известно.

Таким образом, весь предлагаемый метод анализирует вначале признаки вокселей, определяя для каждого вокселя его принадлежность к тому или иному супервокселю при помощи эвристического алгоритма сегментации VCCS, а затем, используя нейросетевую архитектуру из комбинации сетей MiniPointNet и MiniDGCNN, выделяет векторы признаков, соответствующие каждому супервокселю. Полученное таким образом распределение супервокселей в пространстве признаков может быть подвергнуто кластеризации классическими алгоритмами, формируя конечные кластеры точек неизвестных объектов.

В процессе реализации метода были использованы открытые библиотеки Numpy, Open3D, Scikit Learn, PyTorch для языка Python и библиотека PCL для языка C++.

#### Обучение нейронной сети

Для достижения лучших обобщающих способностей нейронной сети и предотвращения переобучения требуется обучающая выборка с объектами разной формы и текстуры. Для создания обучающей выборки использовался набор объектов из открытых наборов данных T-Less [13], ModelNet40 [14], YCB [15] и ряд других моделей. Количество моделей обучающей выборки составило 135, количество моделей неизвестных объектов составило 7. Примеры облаков точек, относящихся к обучающей выборке, тестовой выборке с известными объектами и тестовой выборке с неизвестными объектами, приведены на рис. 3. Облака точек были получены путем расстановки 12 RGBD-камер по полуокружности в среде виртуального моделирования.

Для решения проблемы отставания программно-созданного датасета от реальных данных, были предприняты следующие меры:

- К каждому пикселю изображения глубины, получаемому с RGBD-камеры, прибавлялась случайная ошибка, соответствующая нормальному распределению с математическим ожиданием  $\mu = 0$  м и стандартным отклонением  $\sigma = 0.002$  м;
  - При формировании каждого облака точек использовалось случайное освещение виртуальной сцены.
- Для формирования обучающей выборки также была внесена модификация в алгоритм VCCS, при которой в

механизме кластеризации учитывались истинные метки вокселей. Пример результатов работы алгоритма VCCS без использования информации об истинных сегментах объектов и с ней приведен на рис. 4.

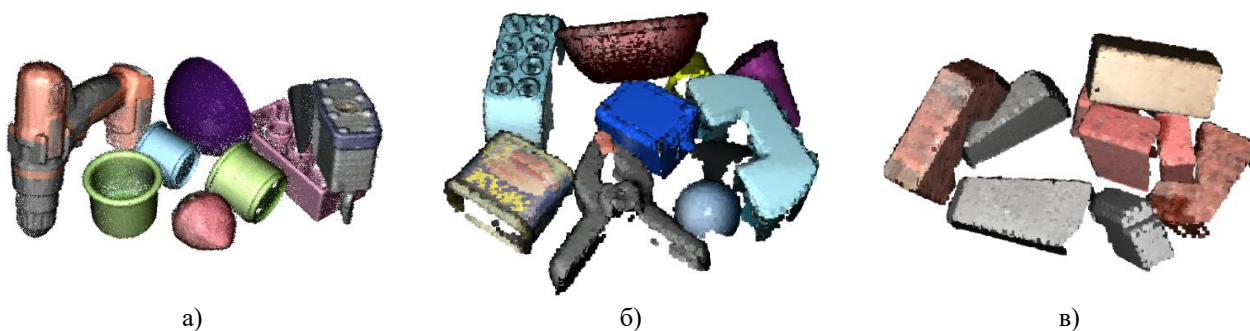


Рисунок 3. Примеры сцен: а) из обучающей выборки; б) из тестового множества сцен с известными объектами; в) из тестового множества сцен с неизвестными объектами

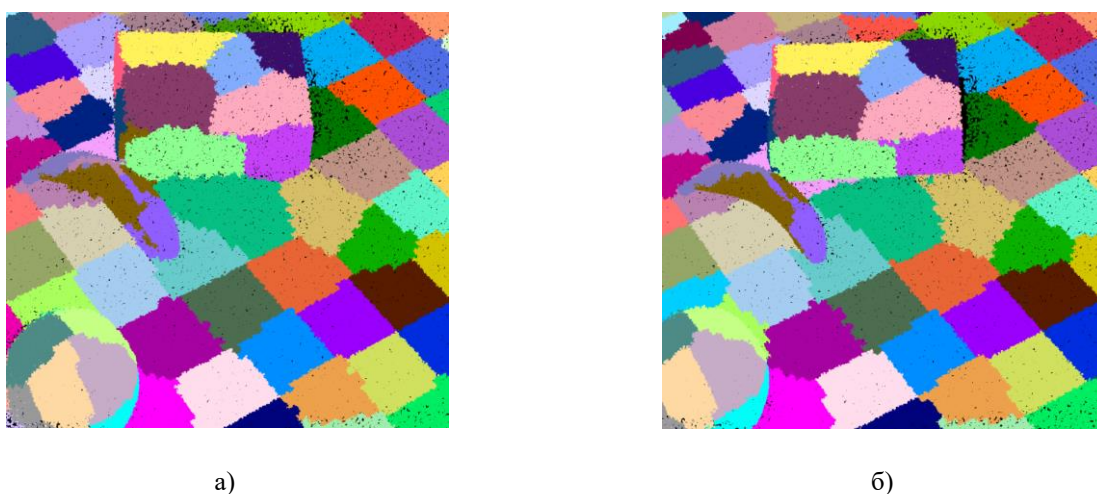


Рисунок 4. Сегментация алгоритмом VCCS: а) сегментации VCCS без использования информации об истинных сегментах объектов; б) сегментация VCCS с использованием информации об истинных сегментах объектов

Аугментация данных является важным элементом при обучении нейронных сетей. Для расширения обучающей выборки в данной работе использовалась аугментация данных путем формирования нескольких выборок из  $n = 160$  супервокселей исходного набора супервокселей. Полученные после аугментации данные состояли из 3 выборок: 360 элементов обучающей выборки с известными объектами, 40 элементов тестовой выборки с известными объектами и 40 элементов тестовой выборки с неизвестными объектами. Для того чтобы нейросеть обучалась работать с облаками точек независимо от их ориентации и для дополнительного расширения обучающей выборки, использовалась также динамическая аугментация данных, при которой набор супервокселей перед подачей на вход нейронной сети, поворачивался на случайно выбранный угол вокруг оси  $Z$  и отражался относительно плоскостей  $XOZ$  и  $YOZ$ .

Разработка функции потерь для тренировки нейронной сети – крайне важный этап разработки нейросетевых моделей. Правильный выбор функции способствует тому, что нейронная сеть обучается извлекать нужные признаки из исходной выборки данных и правильно распределять их в пространстве признаков. Для обучения сети была использована функция потерь, аналогичная [16]. Данная функция потерь отличается от предложенной в [16] тем, что использует  $L_2$ -норму векторов, весовой коэффициент  $\alpha = 1.5$  и не использует возведение в квадрат:

$$Loss = \alpha \cdot Loss_I + Loss_C, \quad (8)$$

где  $\alpha$  – весовой коэффициент ошибки  $Loss_I$ .

С помощью функции  $Loss_I$  рассчитывается среднее расстояние между каждым вектором признаков супервокселей  $i$ -го объекта и их центроидом:

$$Loss_I = \frac{1}{I} \sum_{i=1}^I \frac{1}{N_i} \sum_{j=1}^{N_i} \max(0, \|\mu_i - e_j\| - \delta_1), \quad (9)$$

$I$  – количество объектов в рассматриваемом наборе супервокселей,  
 $N_i$  – количество супервокселей  $i$ -го объекта,  
 $\mu_i$  – координаты центроида кластера  $i$ -го объекта в пространстве признаков,  
 $e_j$  – координаты  $j$ -го супервокселя  $i$ -го объекта в пространстве признаков,  
 $\delta_1$  – слагаемое, отражающее максимальное допустимое расстояние, которое не будет оказывать влияния на величину функции потерь (использовано значение  $\delta_1 = 0.5$ ).

С помощью функции  $Loss_C$  рассчитывается среднее расстояние между центроидами кластеров векторов признаков супервокселей объектов:

$$Loss_C = \frac{1}{I(I-1)} \sum_{i_A=1}^I \frac{1}{N_i} \sum_{i_B=1}^I \max(0, \delta_2 - \|\mu_{i_A} - \mu_{i_B}\|), i_A \neq i_B \quad (10)$$

$I$  – количество объектов в рассматриваемом наборе супервокселей,  
 $\mu_{i_A}$  – координаты центроида кластера объекта  $i_A$  в пространстве признаков,  
 $\mu_{i_B}$  – координаты центроида кластера объекта  $i_B$  в пространстве признаков,  
 $N_i$  – количество супервокселей  $i$ -го объекта,  
 $\delta_2$  – слагаемое, определяющее величину допустимого минимального расстояния между центроидами, которое не будет оказывать влияния на величину функции потерь (использовано значение  $\delta_2 = 3$ ).

Таким образом, представленная функция потерь стимулирует нейронную сеть выдавать признаки супервокселей одного объекта, расположенные близко друг к другу в пространстве признаков, и отталкивать кластеры векторов признаков супервокселей разных объектов друг от друга. Стоит отметить, что ключевую роль играют не значения векторов признаков на выходе нейронной сети, а кластеры, образованные ими. Иными словами, нейронная сеть обучается правильному размещению точек в пространстве признаков. Аналогичный подход к задаче сегментации объектов облака точек можно встретить в [16].

Для обучения нейронной сети использовалась открытая библиотека PyTorch. Используя оптимизатор SGD, коэффициент скорости обучения 0.001, момент обучения 0.9 и размер пакета 2, нейронная сеть обучалась в течение 328 эпох. На рис. 5 представлен график процесса обучения нейронной сети. Как видно из графика, значение функции потерь для всех выборок снижалось с ростом количества эпох.

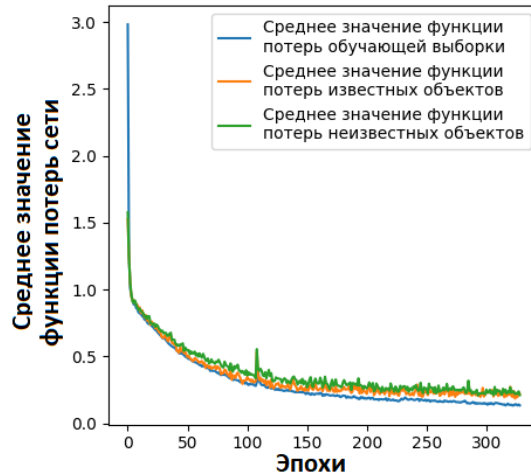


Рисунок 5. Среднее значение функции потерь сети для 3 множеств данных

### Оценка метода

Поскольку предлагаемый метод производит сегментацию облака точек на экземпляры объектов, оценка метода производилась с помощью широко известного коэффициента Жаккара, рассчитанного для каждого истинного экземпляра и отражающего отношение пересечения истинного и предсказанного сегментов к их объединению:

$$IoU = \frac{1}{N} \sum_i^N \frac{g_i \cap r_j}{g_i \cup r_j}, \quad (11)$$

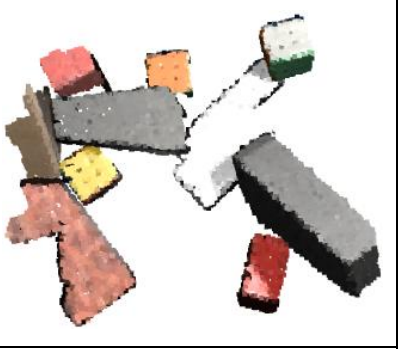
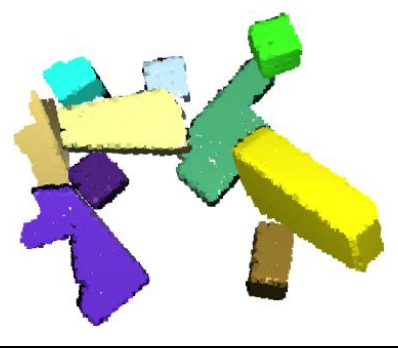
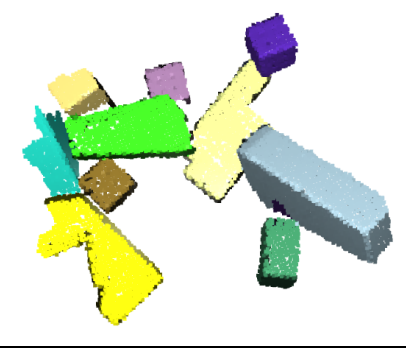
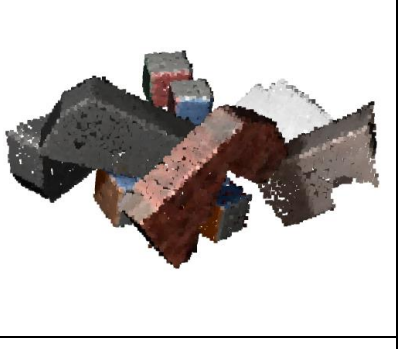
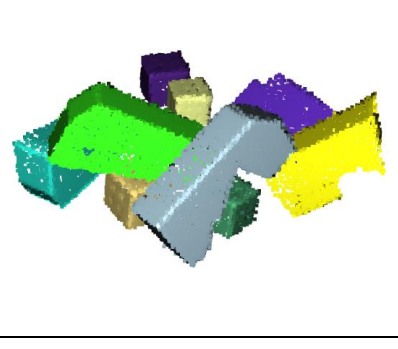
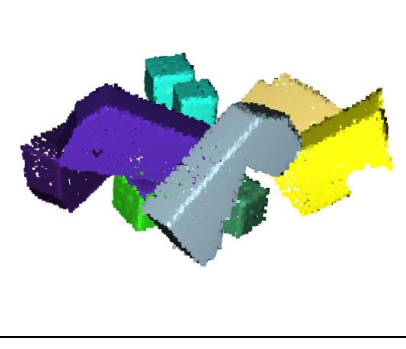

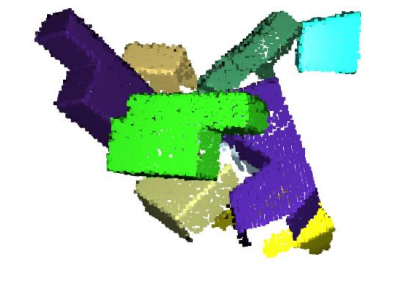
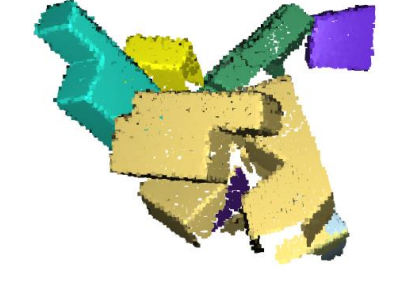
где  $N$  – количество объектов в множестве супервокселей;

$r_j$  – предсказанный сегмент, который имеет наибольшее количество общих точек с истинным сегментом  $g_i$ , аналогично [4].



Результаты работы метода на неизвестных объектах, не участвовавших при обучении сети, при разных коэффициентах Жаккара представлены в табл. 1. Как видно из таблицы, чем выше полученный коэффициент Жаккара, тем лучше качество сегментации. Экспериментальные исследования показали, что средний коэффициент Жаккара для элементов обучающей выборки составляет 0.82, для элементов тестовой выборки с известными объектами - 0.812, для тестовой выборки с неизвестными объектами - 0.804.

Таблица 1. Результаты сегментации неизвестных объектов

IoU	Облако точек	Истинная сегментация	Предсказанная сегментация
0.97			
0.86			
0.58			

На рис. 6 представлена круговая диаграмма, отражающая средние временные затраты реализованного метода для одного исходного облака точек. Расчеты производились на основе набора облаков точек со средним количеством точек  $n = 271920$  и размерностью вектора признаков точек  $D = 9$ . Как видно из диаграммы, большую часть времени (0.295 с) занимает работа алгоритма сегментации VCCS и алгоритма выборки 160 супервокселей (0.209 с). Расчет нейронной сети занимает 0.07 с и кластеризация полученных данных с помощью алгоритма DBSCAN занимает 0.005 с. Запуск производился на персональном компьютере с процессором AMD Ryzen 5 3500U и мобильным графическим адаптером Nvidia GTX 1050.

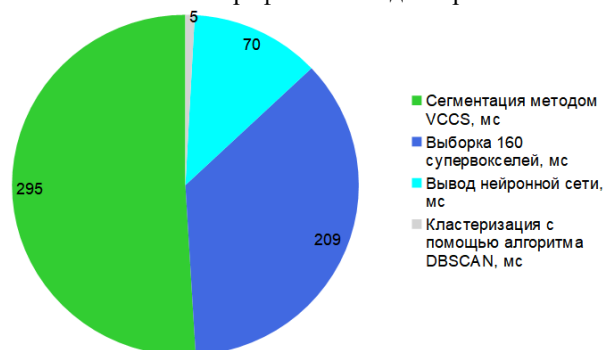


Рисунок 6. Круговая диаграмма затрат работы метода

Таким образом, проведенные исследования позволяют судить, что предложенный метод может быть использован для сегментации неизвестных объектов в облаке точек. Среди достоинств метода можно выделить:

1. Способность метода работать с произвольными объектами, анализируя их признаки;
2. Комбинация эвристического подхода и подхода, основанного на машинном обучении, дающая более высокую точность сегментации трехмерных объектов;
3. Благодаря применению алгоритма VCCS, на вход нейронной сети подается набор супервокселей с устраненными шумами и равномерной плотностью точек в них, что повышает качество работы нейронных сетей.

#### **Заключение**

В статье была рассмотрена задача сегментации облака точек, содержащего априорно неизвестные объекты. На основе обзора существующих задач манипуляционной робототехники и актуальных методов сегментации данных для систем восприятия робота был разработан и представлен метод сегментации облака точек сцены. Метод оперирует набором супервокселей априорно неизвестных объектов на основе использования эвристического метода сегментации VCCS и нейронной сети, основанной на архитектурах нейронных сетей PointNet и DGCNN. Результатом обработки являются векторы признаков, позволяющие с достаточной надежностью различить экземпляры присутствующих в сцене объектов.

Проведенные исследования разработанного метода показали, что средний коэффициент Жаккара на выборке облаков точек с априорно неизвестными объектами составляет  $IoU = 0.804$ . Дальнейшими направлениями работы будут исследования оптимальной архитектуры сетей комплекса, функции потерь и разработка расширенной обучающей выборки, совмещающей в себе данные, полученные при виртуальном моделировании и данные с реальных RGBD-камер.

#### **Список литературы**

1. Лесков А. Г., Морошкин С. Д. Система технического зрения для определения расположения объектов // Экстремальная робототехника. Труды Международной научно-технической конференции – 2016 – Т. 1, № 1, с. 287-291;
2. Lei Q., Meijer J., Wisse M. A survey of unknown object grasping and our fast grasping algorithm-C shape grasping // 2017 3rd International Conference on Control, Automation and Robotics (ICCAR) – Nagoya, Japan, 2017 – pp. 150-157;
3. Sauvet B., Levesque F., Park S., Cardou P. Model-based grasping of unknown objects from a random pile // Robotics, 2019 – vol. 8, no. 3, 18 p.;
4. Verdoja F., Thomas D., Sugimoto A. Fast 3D point cloud segmentation using supervoxels with geometry and color for 3D scene understanding. // IEEE International Conference on Multimedia and Expo. – Hong Kong, 2017 – pp. 1285–1290;
5. Papon J., Abramov A., Schoeler M., Worgotter F. Voxel cloud connectivity segmentation-supervoxels for point clouds // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Portland, USA, June 2013). – 2013. – pp. 2027-2034;
6. Danielczuk M., Matl M., Gupta S., Li A., Lee A., Mahler J., Goldberg K. Segmenting unknown 3D objects from real depth images using Mask R-CNN trained on synthetic data // 2019 International Conference on Robotics and Automation (ICRA) – 2019 – pp. 7283-7290;
7. Chen J., Kira Z., Cho. Y. K. LRGNet: Learnable Region Growing for Class-Agnostic Point Cloud Segmentation // IEEE Robotics and Automation Letters – 2021 – Vol. 6, no. 2, pp. 2799-2806;
8. Xiang Y., Xie C., Mousavian A., Fox D. Learning RGB-D feature embeddings for unseen object instance segmentation // 4th Conference on Robot Learning (CoRL 2020) – 2020 – pp. 461-470;
9. Charles R. Qi., Hao S., Kaichun M., Leonidas J. G. PointNet: deep learning on point sets for 3D classification and segmentation // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – 2017 – pp. 652-660;
10. Charles R. Qi., Hao S., Kaichun M., Leonidas J. G. Pointnet++: Deep hierarchical feature learning on point sets in a metric space // Neural Information Processing Systems (4-9 December 2017, Long Beach, CA, USA) – 2017 – pp. 5099-5108;
11. Wang Y., Sun Y., Liu Z., Sarma S.E., Bronstein M.M. Solomon J.M. Dynamic Graph CNN for Learning on Point Clouds // ACM Trans. Graphics (TOG) – 2019 – Vol. 38, no. 5. pp. 1-12;
12. Luo Z., Xie Z., Wan J., Zeng Z., Liu L., Tao L. Indoor 3D Point Cloud Segmentation Based on Multi-Constraint Graph Clustering // Remote Sensing – Vol. 15, no. 1 – 25 p.;
13. Hodan T., Haluza P., Obdrzalek S., Matas J., Lourakis M., Zabulis X. T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects // IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, Santa Rosa, USA – 2017 – pp. 880-888;

14. Wu Z., Song S., Khosla A., Yu F., Zhang L., Tang X., Xiao J. 3D ShapeNets: A Deep Representation for Volumetric Shapes // 2015 IEEE Conference on Computer Vision and Pattern Recognition (Boston, MA, USA) – pp. 1912-1920;
15. Calli B., Singh A., Walsman A., Srinivasa S., Abbeel P., Dollar A. The YCB Object and Model Set: Towards Common Benchmarks for Manipulation Research // 2015 IEEE International Conference on Advanced Robotics (Istanbul, Turkey) – 2015 – pp. 510-517;
16. Wang X., Liu S., Shen X., Shen C., Jia J. Associatively Segmenting Instances and Semantics in Point Clouds // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Long Beach, CA, USA) – 2019 – pp. 4091-4100.

---

## References

---

1. Leskov A. G., Moroshkin S. D. Technical vision system for determining the location of objects. Extreme robotics. Proceedings of the International Scientific and Technical Conference – 2016 – Vol 1, no. 1, pp. 287-291;
2. Lei Q., Meijer J., Wisse M. A survey of unknown object grasping and our fast grasping algorithm-C shape grasping // 2017 3rd International Conference on Control, Automation and Robotics (ICCAR) – Nagoya, Japan, 2017 – pp. 150-157;
3. Sauvet B., Levesque F., Park S., Cardou P. Model-based grasping of unknown objects from a random pile // Robotics, 2019 – vol. 8, no. 3, 18 p.;
4. Verdoja F., Thomas D., Sugimoto A. Fast 3D point cloud segmentation using supervoxels with geometry and color for 3D scene understanding. // IEEE International Conference on Multimedia and Expo. – Hong Kong, 2017 – pp. 1285–1290;
5. Papon J., Abramov A., Schoeler M., Worgotter F. Voxel cloud connectivity segmentation-supervoxels for point clouds // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), (Portland, USA, June 2013). – 2013. – pp. 2027-2034;
6. Danielczuk M., Matl M., Gupta S., Li A., Lee A., Mahler J., Goldberg K. Segmenting unknown 3D objects from real depth images using Mask R-CNN trained on synthetic data // 2019 International Conference on Robotics and Automation (ICRA) – 2019 – pp. 7283-7290;
7. Chen J., Kira Z., Cho. Y. K. LRGNet: Learnable Region Growing for Class-Agnostic Point Cloud Segmentation // IEEE Robotics and Automation Letters – 2021 – Vol. 6, no. 2, pp. 2799-2806;
8. Xiang Y., Xie C., Mousavian A., Fox D. Learning RGB-D feature embeddings for unseen object instance segmentation // 4th Conference on Robot Learning (CoRL 2020) – 2020 – pp. 461-470;
9. Charles R. Qi., Hao S., Kaichun M., Leonidas J. G. PointNet: deep learning on point sets for 3D classification and segmentation // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) – 2017 – pp. 652-660;
10. Charles R. Qi., Hao S., Kaichun M., Leonidas J. G. Pointnet++: Deep hierarchical feature learning on point sets in a metric space // Neural Information Processing Systems (4-9 December 2017, Long Beach, CA, USA) – 2017 – pp. 5099-5108;
11. Wang Y., Sun Y., Liu Z., Sarma S.E., Bronstein M.M. Solomon J.M. Dynamic Graph CNN for Learning on Point Clouds // ACM Trans. Graphics (TOG) – 2019 – Vol. 38, no. 5. pp. 1-12;
12. Luo Z., Xie Z., Wan J., Zeng Z., Liu L., Tao L. Indoor 3D Point Cloud Segmentation Based on Multi-Constraint Graph Clustering // Remote Sensing – Vol. 15, no. 1 – 25 p.;
13. Hodan T., Haluza P., Obdrzalek S., Matas J., Lourakis M., Zabulis X. T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects // IEEE Winter Conference on Applications of Computer Vision (WACV), 2017, Santa Rosa, USA – 2017 – pp. 880-888;
14. Wu Z., Song S., Khosla A., Yu F., Zhang L., Tang X., Xiao J. 3D ShapeNets: A Deep Representation for Volumetric Shapes // 2015 IEEE Conference on Computer Vision and Pattern Recognition (Boston, MA, USA) – pp. 1912-1920;
15. Calli B., Singh A., Walsman A., Srinivasa S., Abbeel P., Dollar A. The YCB Object and Model Set: Towards Common Benchmarks for Manipulation Research // 2015 IEEE International Conference on Advanced Robotics (Istanbul, Turkey) – 2015 – pp. 510-517;
16. Wang X., Liu S., Shen X., Shen C., Jia J. Associatively Segmenting Instances and Semantics in Point Clouds // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (Long Beach, CA, USA) – 2019 – pp. 4091-4100.