

СЕГМЕНТАЦИЯ 3D МОДЕЛЕЙ ДАННЫХ С ПОМОЩЬЮ МУЛЬТИМОДАЛЬНОГО ДИНАМИЧЕСКОГО ГРАФА CNN

© 2024 А.В. Вохминцев^{1,2}, В.Р. Аббазов¹, М.А. Романов¹

¹ Челябинский государственный университет
(454001 Челябинск, ул. Братъев Кашириных, д. 129),

² Югорский государственный университет
(628012 Ханты-Мансийск, ул. Чехова, д. 16)

E-mail: vav@csu.ru, std.romanov.ma@gmail.com, abbavar@yandex.ru

Поступила в редакцию: 01.04.2024

В работе предложен метод семантической сегментации облаков точек в виде рельефа местности с использованием мультимодальной архитектуры сверточной нейронной сети на основе регулярного динамического взвешенного графа, которая позволяет получать точное решение задачи семантической сегментации, используя комбинацию геометрических и цветовых признаков точек. Метод может быть эффективно использован для разреженных, зашумленных, неоднородных и невыпуклых облаков точек. В работе было проведено компьютерное моделирование известных методов для семантической сегментации 3D данных с использованием эталонной коллекции данных ModelNet 40 и набора данных археологических памятников бронзового века Южного Зауралья, а именно данных, полученных в результате тахеометрической съемки комплекса археологических памятников в долине реки Синташта с использованием тахеометра Trimble 3300. Был проведен сравнительный анализ предложенного метода и современных методов 3D семантической сегментации с разными комбинациями входных признаков облаков точек, также в работе исследовано влияние на точность семантической сегментации способа формирования облака точек: в первом случае исследовалось облако точек из эталонного набора данных во втором случае применены варианты с использованием 3D регистрации на основе алгоритма ICP (iterative closest point).

Ключевые слова: сегментация 3D объектов, graph convolutional neural networks, регистрация облаков точек.

ОБРАЗЕЦ ЦИТИРОВАНИЯ

Вохминцев А.В., Аббазов В.Р., Романов М.А. Сегментация 3D моделей данных с помощью мультимодального динамического графа CNN // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. 2024. Т. 13, № 2. С. 23–38. DOI: 10.14529/cmse240202.

Введение

При дешифрировании археологических памятников исследователи часто используют методы классификации и сегментации 3D данных на основе различных архитектур сверточных нейронных сетей. Методы можно поделить на не прямые, например, мультимодальные сверточные нейронные сети (MVCNN, multi-view convolutional neural network) [1] и прямые: PointNet [2], PointNet++ [3], на основе графа сверточной нейронной сети (GCNN, Graph Convolutional Neural Networks) [4]. На основе GCNN разработаны различные модификации: на основе динамического графа сверточной нейронной сети (DGCNN, dynamic graph CNN) [5], регулярного графа сверточной нейронной сети (RGCNN, regularized graph CNN) [6], ConvPoint [7]. Применение не прямых методов ограничено, так как они обеспечивают хорошее качество семантической обработки только для простых полигональных моделей 3D данных, не прямые методы имеют низкую производительность и часто требуются к памяти для хранения результатов. PointNet и PointNet++ основаны на локальной обработке точек, не используют геометрические отношения между точками и инвариантны

к перестановкам, данные методы находят широкое применение в промышленном дизайне. Модификации прямых методов на основе GCNN при семантической обработке 3D данных используют информацию о форме и поверхности объекта, процесс классификации и сегментации в них основан на выполнении множественных операций фильтрации и свертки сигналов на пространственных динамических графах в спектральной области. Например, в методе на основе RGGNN для описания связности компонент пространственного графа используется матрица Кирхгофа, построение которой имеет большую вычислительную сложность $O(n^3)$, что ограничивает применение метода при анализе крупномасштабных 3D сцен. GCNN и DGCNN плохо работают с локальными признаками объектов и находят основное применение при семантической классификации 3D объектов. Главный недостаток DGCNN и RGCNN связан с размерностью анализируемых облаков точек: методы сегментации на основе этих сетей хорошо подходят для семантической обработки эталонных коллекций 3D объектов, но плохо применимы к обработке реальных данных в виде плотных крупномасштабных облаков точек. Точность процедуры сегментации на основе различных модификаций GCNN зависит от формы объекта и способа формирования облака точек: хорошо сегментируются только исходные или идеально выровненные облака точек.

В последнее время был предложен ряд эффективных архитектур нейронных сетей для решения задачи классификации и семантической сегментации 3D данных, таких как, ConvPoint [7], KPConv [8], ShellNet [9] и Superpoint Transformer [10]. В табл. 1 представлены количественные результаты, которые иллюстрируют современное состояние методов для решения задачи 3D семантической сегментации с использованием коллекции данных Dales (A Large-scale Aerial LiDAR Data Set for Semantic Segmentation) [11]. В исследовании [7] предложен оригинальный метод, который использует непрерывные свертки при обработке облаков точек с неструктурированной природой и оказывается более эффективным, чем классические методы для 3D семантической сегментации [4–6]. ConvPoint в отличие от других методов способен работать с крупномасштабными наборами данных, которые могут включать тысячи плотных облаков точек большой размерности. Данная особенность и хорошие показатели качества по метрикам общей точности (OA, Overall Accuracy) и среднего пересечения над объединением (mIoU, mean intersection over union) делают данный метод кандидатом для решения поставленной в работе задачи по 3D семантической сегментации археологических памятников бронзового века. В исследовании [8] представлен новый подход к свертке точек KPConv для обработки облаков точек, который отличается деформируемой операцией свертки. Предложенный метод хорошо адаптируется к локальной геометрии облака точек и эффективен для обработки облаков точек различной плотности. KPConv был апробирован на многих наборах данных, кроме Dales, и для всех коллекций показывает хорошие показатели в терминах метрик OA и mIoU.

Методы PointNet и PointNet++ были одними из первых методов для 3D семантической сегментации, тем не менее они по-прежнему занимают высокие места в рейтинге качества, а для многих коллекций данных они показывают наилучшие результаты. В методе PointNet++ используется нейронная сеть с иерархической структурой для обработки наборов точек в метрическом пространстве. Данный метод использует метрику расстояния базового пространства и эффективно фиксирует локальные объекты в нескольких масштабах. PointNet++ показывает выдающиеся результаты для наборов данных с неоднородной плотностью выборки, а также отличается высокой производительностью.

В исследовании [9] предложена эффективная сверточная нейронная сеть ShellNet, использующая статистические данные из концентрических сферических оболочек для определения репрезентативных признаков в облаке точек при вычислении свертки. Предложенный подход позволяет с одной стороны существенно увеличить скорость обучения нейронной сети, а с другой стороны позволяет достигать самых современных результатов при классификации 3D объектов, сегментации частей объектов и семантической сегментации 3D сцен. Архитектура ShellNet подчеркивает потенциал оптимизированных нейронных сетей в обработке контекстуально сложных и крупномасштабных 3D сцен.

В исследовании Superpoint Transformer [10] представлен новый подход для эффективной семантической сегментации крупномасштабных 3D сцен на основе архитектуры трансформера. В архитектуре используется быстрый алгоритм для разбиения облаков точек на иерархические структуры суперпунктов, что позволяет значительно ускорить процедуру предварительной обработки 3D данных. В Superpoint Transformer используется разреженный механизм самовнимания для понимания взаимосвязей между суперпунктами в нескольких масштабах. Superpoint Transformer показывает высокое качество 3D семантической сегментации на тестовых наборах данных, обладает высокой производительностью, позволяет в компактной форме хранить модели в памяти.

Таблица 1. Результаты 3D семантической сегментации на коллекции данных Dales

Рейтинг	Модель	mIoU	Общая точность	Размер модели	Год
1	KPConv	81.1	97.8	14.1M	2019
2	Superpoint Transformer	79.6	97.5	212K	2023
3	SuperCluster	77.3	97.2	210M	2024
4	PointNet++	68.3	95.7	3.0M	2017
5	ConvPoint	67.4	97.2	4.7M	2018
6	SPG	60.6	95.5	280K	2018
7	PointCNN	58.4	97.2	N/A	2018
8	ShellNet	57.4	96.4	N/A	2019

Данная статья посвящена решению задачи повышения точности методов 3D сегментации облаков точек, в ней предложен метод сегментации на основе мультимодального динамического взвешенного графа DGCNN*, который использует лучшие идеи DGCNN и RGCNN, но свободен от ряда их ключевых недостатков. В этой статье мы исследуем методы сегментации, основанные на построении динамического графа CNN, методы на основе архитектуры трансформера не рассматриваются в данной статье.

В разделе 1 представлена архитектура мультимодального динамического графа DGCNN*, а также рассмотрены вопросы предварительной обработки облаков точек и их регистрации при построении 3D с разных углов обзора на сцене. В разделе 2 представлен алгоритм семантической сегментации 3D данных на основе DGCNN*. В разделе 3 представлены результаты компьютерного моделирования при решении задачи исследования структуры археологических данных на примере созданного набора данных, содержащего археологические памятники бронзового века на территории Южного Зауралья. В заключении представлены результаты применения алгоритма сегментации на основе DGCNN* на примере набора данных ModelNet 40, указано направление дальнейших исследований.

1. Мультимодальный динамический граф DGCNN*

1.1. Определение динамического графа сверточной нейронной сети DGCNN*

Археологические памятники включают в себя множество объектов и артефактов, поэтому метод сегментации должен извлекать как глобальные признаки памятника, так и локальные признаки, которые связаны с конкретными объектами или их частями. Входные данные для DGCNN* представляют плотное нерегулярное облако точек $C = \{c_1, \dots, c_n\}$ в \mathbb{R}^3 , где $i = 1, \dots, n$, $c_i = (c_i^x, c_i^y, c_i^z, n_i^x, n_i^y, n_i^z, R_i, G_i, B_i)^\top$ — вектор признаков точки в виде координат точки $(c_i^x, c_i^y, c_i^z)^\top$, нормалей $(n_i^x, n_i^y, n_i^z)^\top$ и компонент цвета $(R_i, G_i, B_i)^\top$. Пусть m — число семантических меток $L = \{l_1, \dots, l_k\}$ в \mathbb{N} , где $k = 1, \dots, n$, тогда выход сети DGCNN* $C = \{c_1^S, \dots, c_n^S\}$ будет иметь размерность $n \times m$ для каждой точки в облаке C . Каналы цветовой модели тон, насыщенность, значение (HSV, hue, saturation, value) обладают различной типологической информацией, поэтому их можно рассматривать как независимые признаки в отличие от каналов цветовой модели красный, зеленый, синий (RGB, red, green, blue). Поэтому в работе признаки RGB преобразованы в HSV. DGCNN* основана на концепции построения динамического графа, формируемого путем пересчета матрицы Кирхгофа графа в каждом сверточном слое сети. DGCNN* принимает плотные облака точек с выбранным набором признаков, анализирует локальные особенности объектов с использованием множественной операции свертки на графах в специальных слоях EdgeConv, и затем с помощью метрического классификатора на базе двух многослойных сетей прямого распространения (MLP, multi-layer perceptron) и одной радиально-базисной сети (RBF, radial basis function) осуществляет сегментацию 3D объектов.

1.2. Регистрация 3D данных

В археологии дешифрирование археологического памятника осуществляется на основе 3D моделей рельефа, полученных с разных точек обзора. Поэтому возникает необходимость решения задачи реконструкции 3D модели памятника. Введем следующие определения: $X = \{x_1, \dots, x_n\}$ и $Y = \{y_1, \dots, y_m\}$ — исходное и целевое облако точек в \mathbf{R}^3 . Одним из известных решений задачи является итеративный алгоритм ближайших точек (ICP, iterative closest point). На основе ICP авторами работы разработан точный комбинированный алгоритм регистрации 3D данных (FICP, fusion iterative closest point) [12], в котором решение вариационной задачи представлено как

$$J(RV, RD) = \frac{\alpha}{W|A_f|} \sum_{i \in A_f} \|M(RV F_x^i) - M(F_y^i)\|^2 + \frac{(1-\alpha)}{w|A_d|} \sum_{j \in A_d} \|RDx_j + T - y_j\|^2, \quad (1)$$

где F_x^i, F_y^i — особые точки [13]; RV — матрица аффинного преобразования для данных о цветовых признаках кадра; RD и T — матрица поворота и вектор переноса для карты глубины соответственно; M — функция преобразования координат особых точек в систему координат камеры; α, W — весовые коэффициенты; A_f — связи между особыми точками; A_d — связи между соответствующими точками x_j и y_j в облаках. Алгоритм позволяет решить проблему зависимости решения вариационной задачи от правильности выбора начальных значений, используется для точной регистрации облаков точек с произвольным пространственным разрешением и масштабом относительно друг друга.

1.3. Предварительная обработка данных

Большинство 3D датчиков глубины генерируют разреженные, зашумленные и неоднородные облака точек, что оказывает негативное влияние на процесс классификации и сегментации 3D объектов. Поэтому для предобработки в DGCNN* использован алгоритм с контролируемой повышающей дискретизацией облака точек на основе метода k ближайших соседей (kNN), который позволяет получить плотное, полное и однородное облако точек. Функция потерь алгоритма

$$L_{up} = \sum_{i=0}^{N_2} \sum_{j \in K(i)} \mu(\|c_j - c_i\|) \gamma(\|c_j - c_i\|), \quad (2)$$

где N_2 — количество точек в облаке, $K(i)$ — число k соседей для точки c_i , $\mu(r) = r$ — потеря отталкивания, $\gamma(r) = \exp(-\delta r^2)$, δ — гиперпараметр функции быстрого снижения веса γ .

1.4. Построение мультимодального динамического графа

Пусть задан неориентированный граф вида $G = \{P, E, A\}$, где P — множество вершин, E — множество ребер, A — взвешенная симметричная матрица смежности с элементом $(a_{i,j})_{n \times n}$ — вес ребра (i, j) , при этом $a_{i,j} \geq 0$. Тогда из матрицы A можно получить матрицу Кирхгофа с помощью формулы $L_c := D - A$, где D — матрица степеней с элементом $(d_{i,j})_{n \times n} = \sum_{j=1}^n a_{i,j}$. Для удаления вершины графа с большим весов в матрице Кирхгофа необходимо выполнить нормализацию компонентов матрицы $L_c^{\text{sum}} = I - (D^+)^{1/2} A (D^+)^{1/2}$, где D^+ — обратная матрица Мура—Пенроуза, I — единичная матрица, при этом изолированные вершины исключаются из процесса нормализации. Для каждой вершины графа G выполняется установка связи с вектором ее признаков c_i с характеристиками i -ой точки в облаке, данному вектору с помощью функции s_i (сигнал графа) ставится соответствие вида $s \rightarrow \mathbb{R}$. Для создания мультимодального динамического графа G используется алгоритм:

Шаг 1. Определение связей для каждой точки облака точек c_i с другими точками c_j в облаке.

Шаг 2. Определение значений веса ребра

$a_{ij} = \exp -\alpha \left(w_1 \left\| c_i^{xyz} - c_j^{xyz} \right\|^2 + w_2 \left\| n_i^{xyz} - n_j^{xyz} \right\|^2 + w_3 \left\| c_i^{rgb} - c_j^{rgb} \right\|^2 \right)$, где α — гиперпараметр, управляющий балансом между точностью решения и гладкостью графа, w_1, w_2, w_3 — веса групп признаков.

Шаг 3. Стоп.

1.5. Фильтрация и свертка графа

Облако точек представляет собой нерегулярный набор данных с неупорядоченными вершинами c_i , поэтому возникают проблемы с подбором ядра свертки в области вершин. В работе произведена фильтрация векторов признаков c_i в спектральной области с помощью преобразования Фурье графа, а затем применена аппроксимация с помощью многочлена Чебышёва для повышения производительности процедуры фильтрации и свертки графа. Матрица Кирхгофа имеет собственные значения λ_l , такие что $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$, которым соответствуют собственные векторы $\left\{ \mu^{(l)} = \left(\mu_1^{(l)}, \mu_2^{(l)}, \dots, \mu_N^{(l)} \right)^\top \right\}_{l=1}^N$. Собственные векторы образуют ортонормированный базис $U = \left(\mu^{(1)}, \mu^{(2)}, \dots, \mu^{(N)} \right)$. Тогда матрица Кирхгофа $L = U \text{diag}(\lambda) U^\top$, где $\text{diag}(\lambda)$ — диагональная матрица. Прямое преоб-

Разложение Фурье графа определяется по формуле $GF[s](\lambda_l) = \hat{s}(\lambda_l) = \sum_{i=1}^N s(i) (\mu_i^{(l)})^\top$, а обратное — по формуле $IGF[\hat{s}](i) = s(i) = \sum_{l=0}^{N-1} \hat{s}(\lambda_l) \mu_i^{(l)}$. Пусть спектральный фильтр задан вектором $\psi \in \mathbb{R}^N$, тогда компонента вектора ψ_i может быть определена как значение функции $g_\psi : \mathbb{R}_+ \rightarrow \mathbb{R}$ в i -ом собственном числе $\psi_i = g_\psi(\lambda_i)$. Пусть на графе заданы G два сигнала s_1 и s_2 , тогда, используя свойства преобразования Фурье, получим $s_1 * s_2 = IGF[GF([s_1] \cdot GF([s_2]))]$, где \cdot — операция покомпонентного умножения, $*$ — операция свертки. На основе теоремы о свертке функций получим формулы для вычисления свертки двух сигналов на графе

$$(s_1 * s_2)(i) = \sum_{l=0}^{N-1} \hat{s}_1(\lambda_l) \hat{s}_2(\lambda_l) \mu_i^{(l)}, \quad (3)$$

$$s_2 = g_\psi(L_c^{sum}) s_1 = g_\psi(U \text{diag}(\lambda) U^\top) s_1.$$

Для аппроксимации спектральной фильтрации были использованы многочлены Чебышёва, которые позволяют повысить производительность метода сегментации. Пусть N — число коэффициентов Чебышёва в многочлене, тогда запишем функцию g_ψ в виде $g_\psi(x) = \frac{1}{2}x_0 + \sum_{m=1}^{\infty} x_m \bar{T}_m(x), \forall x \in [0, \lambda_N]$. Тогда результат действия n -го многочлена Чебышёва на сигнал s_1 может быть определен по формуле

$$s_2 = g_\psi(L_c^{sum}) s_1 \approx \sum_{m=0}^{M-1} \psi_m T_m(L_c^{sum}) s_1. \quad (4)$$

При $N = 1$ фильтрация с помощью многочлена Чебышёва аналогична работе однослойного персептрона. Известно, что информация о локальных особенностях не теряется при свертке графа с многочленами Чебышёва высокого порядка $N > 3$. Нормализованная матрица Кирхгофа L_c^{sum} разрежена и операция ее умножения на вектор имеет линейную сложность $O(|E|)$. Тогда для первых N степеней многочлена вычислительная сложность операции свертки сигналов s_1 и s_2 с использованием многочлена Чебышёва равна $O(N|E|)$.

1.6. Функция потерь

Функция потерь L_{los} имеет два слагаемых и основана на вычислении мультиклассовой кросс-энтропии L_{CE} , также в нее добавлен параметр регуляризации L_s , связанный с гладкостью сигнала графа s , который вычисляется по трем слоям свертки DGCNN*. Данный параметр позволяет делать объекты смежных вершин более похожими, что облегчает задачу семантической сегментации. Пусть X_i — входной вектор сигнала, Y_i — целевой вектор сигнала, тогда

$$L_{los}(X_i, Y_i) = L_{CE} + L_s = - \sum_{j=1}^M y_{i,j} \log p_{i,j} + \alpha \sum_{k=1}^3 f m_k^\top L_c^{sum} f m_k, \quad (5)$$

где M — количество возможных меток класса; $y_{i,j}$ — бинарный признак (маска класса), $p_{i,j}$ — прогнозируемая вероятность модели, которая определяет вероятность того, что (i, j) маска принадлежит классу j ; $f m_k$ — карта признаков для k слоя в графе свертки; α — гиперпараметр, в работе используем значение равное 10^{-9} . Предложенная функция потерь L_{los} обеспечивает сглаживание Лапласа в пространственной и спектральной областях.

2. Алгоритм сегментации 3D данных

Входными данными алгоритма сегментации на основе DGCNN* является набор данных из 9 элементов для каждой точки: координаты, координаты вектора нормалей и данные о цвете, в процессе работы алгоритма данный набор дополняется нормализованными координатами. Координаты точки позволяют осуществлять глобальное позиционирование точек во всем облаке точек, в то время как нормализованные координаты представляют расположение внутри локального блока данных в облаке точек.

В работе произведена модификация первого сверточного слоя GCNN таким образом, чтобы kNN мог динамически использовать информацию о нормализованных координатах и нормалях векторов при поиске соседей для каждой точки. Далее информация в сети последовательно обрабатывается тремя слоями EdgeConv с коэффициентами Чебышёва, равными 6, 5 и 3 соответственно, двумя слоями MLP и одним слоем с радиально-базисными функциями с максимальным объединением объектов из разных слоев, что позволяет извлекать признаки глобальных и локальных объектов в облаке точек. Каждый слой EdgeConv включает в себя построение мультимодального динамического взвешенного графа, фильтрацию объектов и свертку графа. Эксперименты показали, что для повышения точности сегментации нужно осуществлять пересчет матрицы Кирхгофа графа G для каждого слоя EdgeConv. С помощью слоев EdgeConv и слоя пулинга на выходе нейронной сети получаем исходный набор данных с оценками сегментации для каждого класса. На вход модуля построения графов подаются только нормализованные координаты, а на вход сверточного слоя подаются как исходные, так и нормализованные координаты точек.

Далее представим алгоритм сегментации 3D моделей данных на основе DGCNN*.

procedure DGCNN*

Входные данные: $C = \{c_1, \dots, c_n\}$

Выходные данные: $C = \{c_1^S, \dots, c_n^S\}$

- 1: Инициализация: Чебышёв $\{6,5,3\}$, $k=1$ (Счетчик Чебышёва);
- 2: Дискретизация облака точек (UpSampling);
- 3: Преобразование RGB \rightarrow HSV; Формирование вектора признаков $\{C_i\}$;
- 4: **for** k in $1, \dots, 3$ **do**
- 5: Расчет матрицы Кирхгофа L_k ;
- 6: Нормализация компонентов матрицы L^{sum} ;
- 7: Аппроксимация сигнала графа многочленом Чебышёва g_ψ со степенью k ;
- 8: Свертка графа GCNN: $s_2 = g_\psi(L^{sum})s_1$;
- 9: Формирование признаков динамического графа G ;
- 10: **end**;
- 11: MLP 1 (1024);
- 12: MLP 2 (512);
- 13: concatenation (EdgeConv 2, MLP2);
- 14: Формирование результата сегментации на основе RBF;
- 15: **Вычисление функции потерь** $L_{los}(X_i, Y_i)$ (см. (5)).

На рис. 1 представлена архитектура сверточной нейронной сети на основе регулярного динамического взвешенного графа DGCNN*.

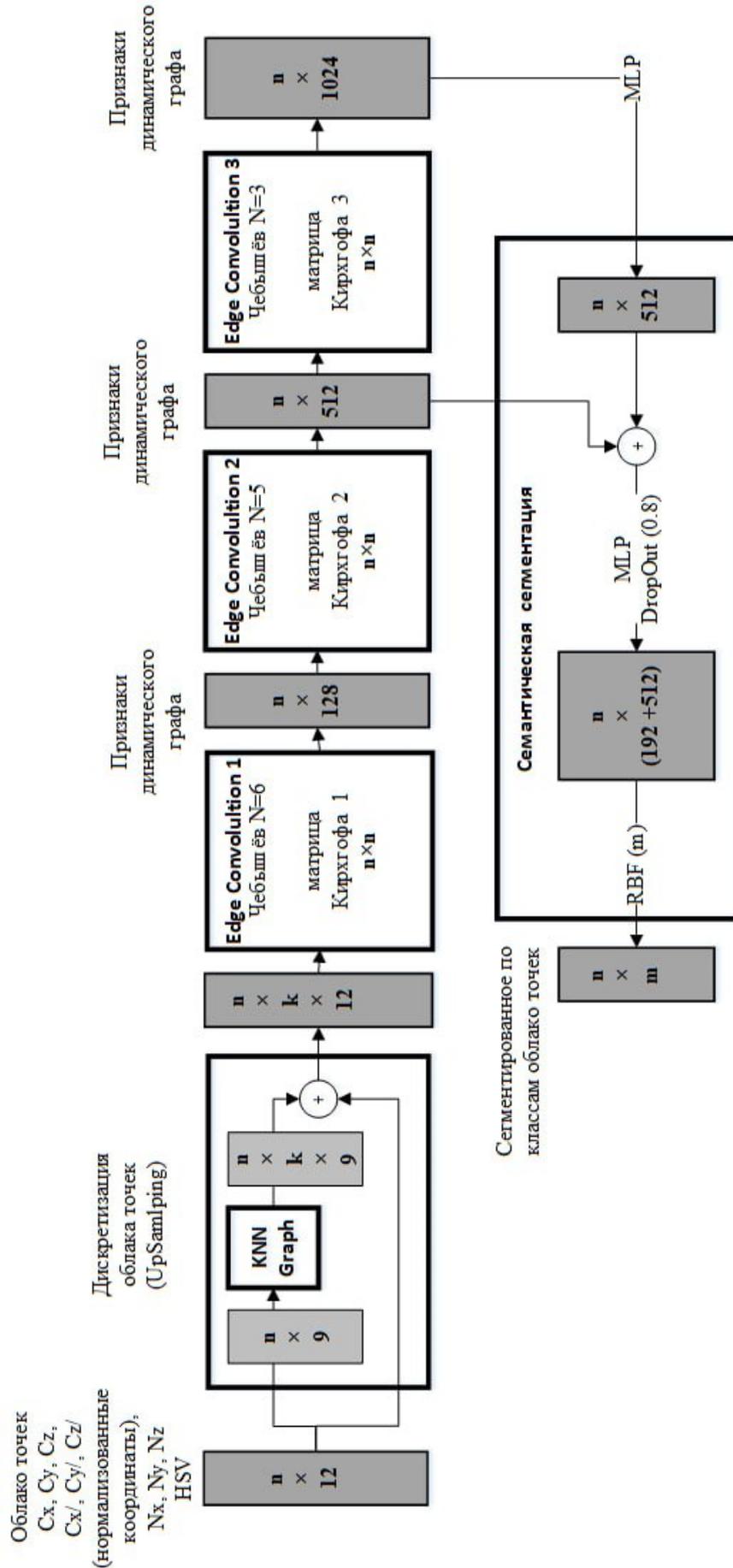


Рис. 1. Архитектура сверточной нейронной сети DGCCNN*

3. Компьютерное моделирование

В работе проведено компьютерное моделирование с использованием эталонного набора ModelNet 40 (табл. 2 и табл. 3) и набора данных, содержащего археологические памятники бронзового века на территории Южного Зауралья [14], а именно облаков точек, полученных в результате тахеометрической съемки археологического комплекса Синташта (табл. 4 и табл. 5) с использованием тахеометра Trimble 3300. Для компьютерного моделирования на коллекции данных, содержащей археологические памятники бронзового века, для регистрации облаков точек были использованы алгоритмы итеративного алгоритма ближайших точек на основе нормалей (NICP, Normal Iterative Closest Point) и комбинированного итеративного алгоритма ближайших точек (FICP).

Таблица 2. Точность 3D сегментации для эталонного набора данных ModelNet 40 для исходного облака точек

Наименование метода	Mean Acc Обучающая выборка	Mean Acc Валид. выборка	Mean Acc Тестовая выборка	F1-мера
Point Net	0.683	0.533	0.520	0.247
Point Net++	0.734	0.648	0.611	0.412
DGCNN	0.891	0.796	0.744	0.747
DGCNN+RGB	0.902	0.841	0.817	0.821
RGCNN	0.896	0.855	0.819	0.789
ConvPoint	0.916	0.903	0.894	0.833
KPCnv	0.941	0.886	0.861	0.866
DGCNN*	0.934	0.923	0.907	0.854

Таблица 3. Точность 3D сегментации для эталонного набора данных ModelNet 40 для выровненного облака точек

Наименование метода	Mean Acc Обучающая выборка	Mean Acc Валид. выборка	Mean Acc Тестовая выборка	F1-мера
DGCNN* + NICP	0.712	0.686	0.674	0.444
DGCNN* + FICP	0.825	0.798	0.776	0.631

Во втором случае облако точек не содержало данных о цвете, поэтому была проведена его раскраска с использованием системы картографирования Trimble MX9 и соответствующих RGB кадров. Из ModelNet 40 было отобрано четыре набора данных с общим количеством объектов 456. Облака точек, полученные с помощью тахеометрической съемки, содержат от 400 до 1200 точек для каждого археологического бронзового памятника, поэтому требуется увеличение плотности облака точек.

Таблица 4. Точность 3D сегментации по данным тахеометрической съемки для исходного облака точек

Наименование метода	Mean Acc Обучающая выборка	Mean Acc Валид. выборка	Mean Acc Тестовая выборка	F1-мера
Point Net	0.623	0.476	0.471	0.256
Point Net++	0.655	0.644	0.602	0.398
DGCNN	0.792	0.713	0.688	0.646
DGCNN+RGB	0.776	0.708	0.653	0.629
ConvPoint	0.833	0.8	0.711	0.625
KPCConv	0.796	0.746	0.735	0.634
RGCNN	0.814	0.776	0.72	0.725
DGCNN*	0.862	0.791	0.773	0.622

Таблица 5. Точность 3D сегментации по данным тахеометрической съемки для выровненного облака точек

Наименование метода	Mean Acc Обучающая выборка	Mean Acc Валид. выборка	Mean Acc Тестовая выборка	F1-мера
DGCNN* + NICP	0.69	0.543	0.511	0.475
DGCNN* + FICP	0.786	0.742	0.753	0.685

Далее представим псевдокод модифицированного иерархического алгоритма агломеративной кластеризации.

procedure GLA*

Входные данные: max distance= 0.3, $P^{IN} = \{p_1^{IN}, \dots, p_n^{IN}\}$

Выходные данные: $P^{OUT} = \{p_1^{OUT}, \dots, p_m^{OUT}\}$

- 1: $points_{norm}$ = Нормализация точек P^{IN} в интервале [0...1];
- 2: $model_{cluster}$ = new Agglomerative Clustering (max distance);
- 3: $model_{cluster}$ = Обучение нейронной сети ($points_{norm}$);
- 4: $Clusters = \{c_1, \dots, c_n\} = model_{cluster}$. Получить кластеры($points_{norm}$);
- 5: **for** Кластер_{*j*} in $model_{cluster}$.Кластеры() **do**
- 6: $points_{cluster}$ = Фильтрация ($points_{norm}$, Кластер_{*j*});
- 7: **if** $points_{cluster}$.Счетчик точек() < p_{aug} **then** // $p_{aug} = 4$, параметр кластеризации
- 8: P^{OUT} .Добавление точек($points_{cluster}$);
- 9: continue;
- 10: **end;**
- 11: $Grid_{points}$ = Создать сетку точек (Крайние точки. $points_{cluster}$);
- 12: $points_{interpolation}$ = Интерполяция ($points_{cluster}$, $Grid_{points}$);
- 13: P^{OUT} .($points_{interpolation}$);
- 14: **end;**
- 15: Формирование результата агломеративной кластеризации.

Для увеличения плотности облака точек и повышения степени равномерности точек в облаке в работе были использованы модифицированный иерархический алгоритм агломеративной кластеризации Agglomerative Clustering, а точнее вариант на основе GLA* (Generic_linkage algorithm) [15] и алгоритм повышающей дискретизации облаков точек на основе графовой сверточной нейронной сети (PU-GCN, point cloud upsampling using graph convolutional networks) [16].

На рис. 2 представлены результаты работы алгоритма GLA* для могильника вблизи п. Осиповка.

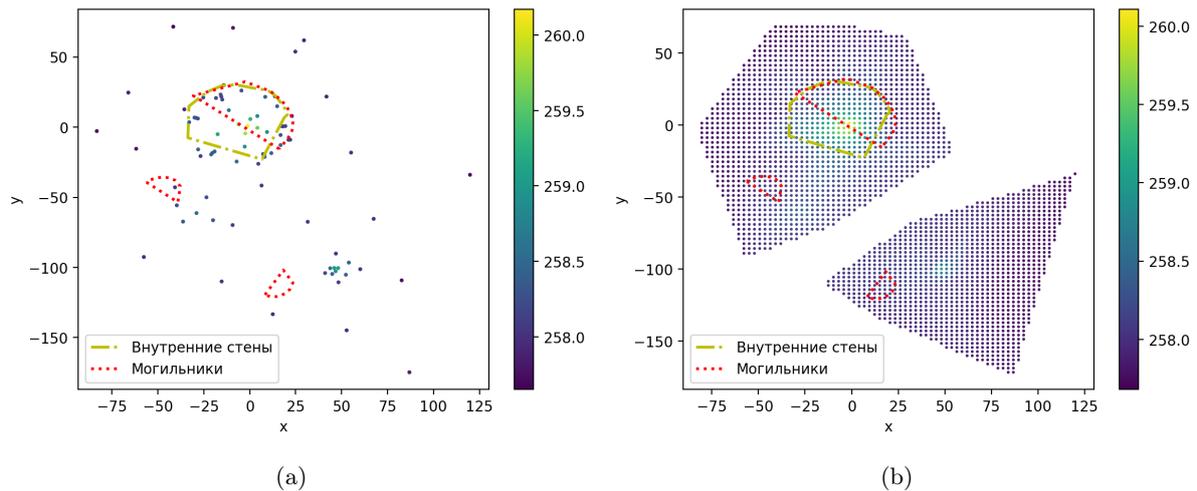


Рис. 2. Результаты повышения размерности облака точек с использованием алгоритма GLA* на примере могильника вблизи п. Осиповка: а) облако точек до применения алгоритма агломеративной кластеризации б) облако точек после применения алгоритма агломеративной кластеризации

На рис. 3 представлены результаты работы алгоритма повышающей дискретизации облаков точек PU-GCN на примере археологического памятника бронзового века вблизи п. Каменка Челябинской области.

Построенные модели обучены на ПК на базе Intel Core i7 с картой Nvidia GeForce GTX 1080Ti в течение 150 эпох. Мы провели дискретизацию облака точек на основе kNN с $k = 20$. Был проведен сравнительный анализ предложенного метода сегментации с известными методами 3D сегментации и разными комбинациями входных признаков облаков точек, исследовано влияние на точность сегментации способа формирования облака точек: в первом случае исследовалось облако точек из эталонного набора данных (табл. 2 и табл. 4), во втором случае применены варианты с использованием 3D регистрации на основе алгоритмов NISР И FICP (табл. 3 и табл. 5).

Моделирование показало превосходство предложенного метода по всем метрикам над Point Net, Point Net++ и DGCNN, для RGCNN и DGCNN+RGB получены близкие по точности результаты по метрике F1 и лучшие результаты по метрике Mean Acc. Использование методов регистрации для формирования облака точек ожидаемо приводит к уменьшению точности процедуры его сегментации, однако метод FICP [17] имеет преимущества по сравнению с NISР.

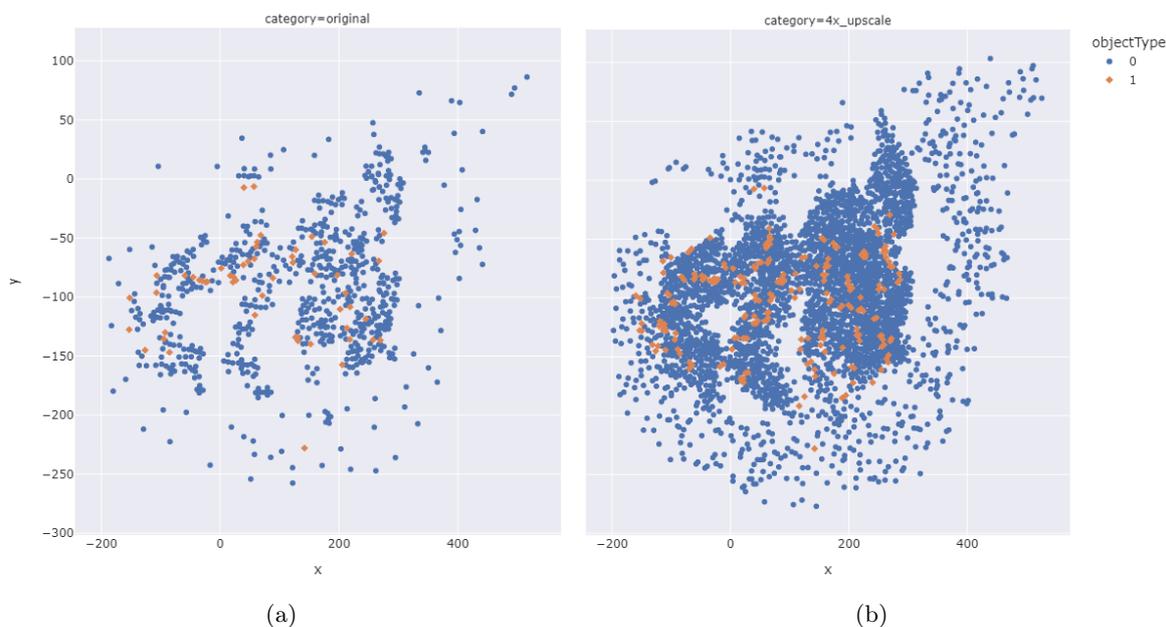


Рис. 3. Результаты повышения плотности облака точек с использованием алгоритма PU-GCN на примере археологического памятника вблизи п. Каменка: а) облако точек до применения алгоритма повышающей дискретизации б) облако точек после применения алгоритма повышающей дискретизации (класс 0 — «фон», класс 1 — «жилищная впадина»)

Заключение

В работе предложен метод семантической сегментации нерегулярных облаков точек на основе мультимодального взвешенного динамического графа DGCNN*. Архитектура сети использует принцип динамического определения соседства точек в облаке на основе данных о геометрии облаков точек и данных о цвете, что позволяет устранить главный недостаток DGCNN и RGCNN, связанный с размерностью обрабатываемых облаков точек. Структура полносвязного слоя в виде метрического классификатора из двух MLP сетей и RFB сети с конкатенацией данных, получаемых с EdgeConv слоев сети, позволяет эффективно обрабатывать локальные и глобальные признаки объектов. Предложенный метод позволяет получать точное решение задачи семантической сегментации для разреженных, зашумленных и неоднородных облаков точек, для 3D сцен с микрорельефом и объектами невыпуклой формы, что важно при обработке геопространственных данных. Метод независим от способа сбора данных: показывает достаточно высокую точность для облаков точек, полученных с использованием вариантов алгоритма ICP.

Исследование выполнено за счет гранта РФФИ (проект 23-11-20007).

Литература

1. Su H., Maji S., Kalogerakis E., Learned-Miller E. Multi-view Convolutional Neural Networks for 3D Shape Recognition // IEEE Proceedings of International Conference on Computer Vision (ICCV) (Santiago, Chile, December 7–13, 2015). P. 945–953. DOI: 10.1109/ICCV.2015.114.

2. Charles R.Q., Su H., Kaichun M., Guibas L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, USA, July 21–26, 2017). P. 77–85. DOI: 10.1109/CVPR.2017.16.
3. Charles R.Q., Li Y., Hao S., Leonidas J.G. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space // Proceedings of 31st Conference on Neural Information Processing Systems (NIPS) (Long Beach, USA, December 4–9, 2017). P. 5099–5108. DOI: 10.48550/arXiv.1706.02413.
4. Zhang Y., Rabbat M. A Graph-CNN for 3D Point Cloud Classification // IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Calgary, AB, Canada, April 15–20, 2018). P. 6279–6283. DOI: 10.1109/ICASSP.2018.8462291.
5. Wang Y., Sun. Y., Liu Z., *et al.* Dynamic Graph CNN for Learning on Point Clouds // ACM Transactions on Graphics. 2019. Vol.38, No. 5. Article 146. P. 1–12. DOI: 10.1145/3326362.
6. Te G., Hu W., Zheng A., Guo Z. RGCNN: Regularized Graph CNN for Point Cloud Segmentation // Proceedings of the 26th ACM international conference on Multimedia (MM '18) (Seoul, Republic of Korea, October 22–26, 2018). ACM, 2018. P. 746–754. DOI: 10.1145/3240508.3240621.
7. Boulch A. ConvPoint: Continuous convolutions for point cloud processing // Computers and Graphics. 2020. Vol. 88. P. 24–34. DOI: 10.1016/j.cag.2020.02.005.
8. Hugues T., Charles R.Q., Deschaud J.E., *et al.* KPConv: Flexible and Deformable Convolution for Point Clouds // IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, Republic of Korea, October 27 – November 2, 2019). P. 6410–6419. DOI: 10.1109/ICCV.2019.00651.
9. Zhang Z., Hua B.S., Yeung S.K. ShellNet: Efficient Point Cloud Convolutional Neural Networks Using Concentric Shells Statistics // IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, Republic of Korea, October 27 – November 2, 2019). P. 1607–1616. DOI: 10.1109/ICCV.2019.00169.
10. Damien R., Hugo R., Loic L. Efficient 3D semantic segmentation with superpoint transformer // IEEE/CVF International Conference on Computer Vision (ICCV) (Paris, France, October 1–6, 2023). P. 17149–17158. DOI: 10.1109/ICCV51070.2023.0157.
11. 3D Semantic Segmentation on DALES. URL: <https://paperswithcode.com/sota/3d-semantic-segmentation-on-dales> (дата обращения: 31.03.2024).
12. Вохминцев А.В., Мельников А.В., Пачганов С.А. Метод навигации и составления карты в трехмерном пространстве на основе комбинированного решения вариационной подзадачи точка-точка ICP для аффинных преобразований // Информатика и ее применения. 2020. Т. 14, № 1. С. 101–112. DOI: 10.14357/19922264200114.
13. Вохминцев А.В., Соченков И.В., Кузнецов В.В., Тихоньких Д.В. Распознавание лиц на основе алгоритма сопоставления изображений с рекурсивным вычислением гистограмм направленных градиентов // Доклады академии наук. 2016. Т. 93, № 1. С. 37–41. DOI: 10.1134/S1064562416010178.
14. Vokhmintcev A.V., Khristodulo O.I, Melnikov A.V., Romanov M.A. Application of Dynamic Graph CNN* and FICP for Detection and Research Archaeology Sites // Analysis of Images,

Social Networks and Texts (AIST 2023). Vol. 14486 / eds. by D.I. Ignatov, *et al.* Cham: Springer, 2024. Lecture Notes in Computer Science. DOI: 10.1007/978-3-031-54534-4_21.

15. Day W.H.E., Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods // *Journal of Classification*. 1984. Vol. 1, no. 1. P. 7–24. DOI: 10.1007/BF01890115.
16. Qian G., Abualshour A., Li G., *et al.* PU-GCN: Point Cloud Upsampling using Graph Convolutional Networks // *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Nashville, TN, USA, June 20–25, 2021)*. DOI: 10.1109/CVPR46437.2021.01151.
17. Vokhmintsev A.V., Khristodulo O.I., Romanov M.A. Semantic Classification and Segmentation of Archaeological Sites Based on a Fusion of Object Detector and 3DEF // *2023 International Russian Automation Conference (RusAutoCon) (Sochi, Russian Federation, September 10–16, 2023)*. P. 122–127. DOI: 10.1109/RusAutoCon58002.2023.10272916.

Вохминцев Александр Владиславович, д.т.н., заведующий, научно-исследовательская лаборатория интеллектуальных информационных технологий и систем, Челябинский государственный университет (Челябинск, Российская Федерация)

Аббазов Валериан Ринатович, лаборант-исследователь, научно-исследовательская лаборатория интеллектуальных информационных технологий и систем, Челябинский государственный университет (Челябинск, Российская Федерация)

Романов Матвей Александрович, лаборант-исследователь, научно-исследовательская лаборатория интеллектуальных информационных технологий и систем, Челябинский государственный университет (Челябинск, Российская Федерация)

DOI: 10.14529/cmse240202

MULTIMODAL DYNAMIC GRAPH CNN FOR 3D SEMANTIC SEGMENTATION

© 2024 A.V. Vokhmintsev^{1,2}, V.R. Abbazov¹, M.A. Romanov¹

¹*Chelyabinsk State University (st. Brat'yev Kashirinykh 129, Chelyabinsk, 454001 Russia),*

²*Yugra Research Institute of Information Technologies
(st. Chekhova 16, Khanty-Mansiysk, 628012 Russia)*

E-mail: vav@csu.ru, std.romanov.ma@gmail.com, abbavar@yandex.ru

Received: 01.04.2024

In this paper, a semantic segmentation method of point clouds in the form of terrain using a new multimodal convolutional neural network architecture based on a regular dynamic weighted graph, which allows to obtain an accurate solution to the segmentation problem based on a fusion of geometric and color features. The method can be effectively used for sparse, noisy, inhomogeneous and non-convex point clouds. The computer modeling of state-of-the-art methods for 3D semantic segmentation was carried out using the reference data collection ModelNet 40 and a data set of archaeological sites of the Bronze Age of the Southern Trans-Urals, namely data obtained as a result of a total station survey (the Trimble 3300 total station) of a complex of archaeological sites in the valley of the Sintashta river. A comparative analysis of the proposed method and state-of-the-art methods for 3D semantic segmentation with different combinations of input features of point clouds was carried out, and the method influence of forming a point cloud on the accuracy of 3D semantic segmentation was also investigated: in the first case, a point cloud from a reference dataset was studied, in the second case, variants using 3D registration based on NICP and FICP algorithms were applied.

Keywords: segmentation of 3D objects, graph convolutional neural networks, point clouds registration.

FOR CITATION

Vokhmintcev A.V., Abbazov V.R., Romanov M.A. Multimodal Dynamic Graph CNN for 3D Semantic Segmentation. Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering. 2024. Vol. 13, no. 2. P. 23–38. (in Russian) DOI: 10.14529/cmse240202.

This paper is distributed under the terms of the Creative Commons Attribution-Non Commercial 4.0 License which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is properly cited.

References

1. Su H., Maji S., Kalogerakis E., Learned-Miller E. Multi-view Convolutional Neural Networks for 3D Shape Recognition. IEEE Proceedings of International Conference on Computer Vision (ICCV) (Santiago, Chile, December 7–13, 2015). P. 945–953. DOI: 10.1109/ICCV.2015.114.
2. Charles R.Q., Su H., Kaichun M., Guibas L.J. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (Honolulu, HI, USA, July 21–26, 2017). P. 77–85. DOI: 10.1109/CVPR.2017.16.
3. Charles R.Q., Li Y., Hao S., Leonidas J.G. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space. Proceedings of 31st Conference on Neural Information Processing Systems (NIPS) (Long Beach, USA, December 4–9, 2017). P. 5099–5108. DOI: 10.48550/arXiv.1706.02413.
4. Zhang Y., Rabbat M. A Graph-CNN for 3D Point Cloud Classification. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (Calgary, AB, Canada, April 15–20, 2018). P. 6279–6283. DOI: 10.1109/ICASSP.2018.8462291.
5. Wang Y., Sun. Y., Liu Z., *et al.* Dynamic Graph CNN for Learning on Point Clouds. ACM Transactions on Graphics. 2019. Vol. 38. no. 5. Article 146. P. 1–12. DOI: 10.1145/3326362.
6. Te G., Hu W., Zheng A., Guo Z. RGCNN: Regularized Graph CNN for Point Cloud Segmentation. Proceedings of the 26th ACM international conference on Multimedia (MM '18) (Seoul, Republic of Korea, October 22–26, 2018). ACM, 2018. P. 746–754. DOI: 10.1145/3240508.3240621.
7. Boulch A. ConvPoint: Continuous convolutions for point cloud processing. Computers and Graphics. 2020. Vol. 88. P. 24–34. DOI: 10.1016/j.cag.2020.02.005.
8. Hugues T., Charles R.Q., Deschaud J.E., *et al.* KPConv: Flexible and Deformable Convolution for Point Clouds. IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, Republic of Korea, October 27 – November 2, 2019). P. 6410–6419. DOI: 10.1109/ICCV.2019.00651.
9. Zhang Z., Hua B.S., Yeung S.K. ShellNet: Efficient Point Cloud Convolutional Neural Networks Using Concentric Shells Statistics. IEEE/CVF International Conference on Computer Vision (ICCV) (Seoul, Republic of Korea, October 27 – November 2, 2019). P. 1607–1616. DOI: 10.1109/ICCV.2019.00169.

10. Damien R., Hugo R., Loic L. Efficient 3D semantic segmentation with superpoint transformer. IEEE/CVF International Conference on Computer Vision (ICCV) (Paris, France, October 1–6, 2023). P. 17149–17158. DOI: 10.1109/ICCV51070.2023.0157.
11. 3D Semantic Segmentation on DALES. URL: <https://paperswithcode.com/sota/3d-semantic-segmentation-on-dales> (accessed: 31.03.2024).
12. Vokhmintsev A.V., Melnikov A.V., Pachganov S.A. Simultaneous localization and mapping method in three-dimensional space based on the combined solution of the point-point variation problem icp for an affine transformation. Informatics and Applications. 2020. Vol. 14, no. 1. P. 101–112. DOI: 10.14357/19922264200114.
13. Vokhmintsev A.V., Sochenkov I.V., Kuznetsov V.V., Tikhonkikh D.V. Face recognition based on matching algorithm with recursive calculation of local oriented gradient histogram. Doklady Mathematics. 2016. Vol. 93, no. 1. P. 37–41. DOI: 10.1134/S1064562416010178.
14. Vokhmintsev A.V., Khristodulo O.I., Melnikov A.V., Romanov M.A. Application of Dynamic Graph CNN* and FICP for Detection and Research Archaeology Sites. Analysis of Images, Social Networks and Texts (AIST 2023). Vol. 14486 / eds. by D.I. Ignatov, *et al.* Cham: Springer, 2024. Lecture Notes in Computer Science. DOI: 10.1007/978-3-031-54534-4_21.
15. Day W.H.E., Edelsbrunner H. Efficient algorithms for agglomerative hierarchical clustering methods. Journal of Classification. 1984. Vol. 1, no. 1. P. 7–24. DOI: 10.1007/BF01890115.
16. Qian G., Abualshour A., Li G., *et al.* PU-GCN: Point Cloud Upsampling using Graph Convolutional Networks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (Nashville, TN, USA, June 20–25, 2021). DOI: 10.1109/CVPR46437.2021.01151.
17. Vokhmintsev A.V., Khristodulo O.I., Romanov M.A. Semantic Classification and Segmentation of Archaeological Sites Based on a Fusion of Object Detector and 3DEF. 2023 International Russian Automation Conference (RusAutoCon) (Sochi, Russian Federation, September 10–16, 2023). P. 122–127. DOI: 10.1109/RusAutoCon58002.2023.10272916.