

Создание полигональных сеток из облаков точек с помощью глубокого обучения для их визуализации в системах смешанной реальности

М.И. Сорокин¹, Д.Д. Жданов¹, А.Д. Жданов¹

¹ Университет ИТМО, Kronverkskiy prospekt, dom. 49, liter A, Санкт-Петербург, 197101, Россия

Аннотация

Системы смешанной реальности создают новые формы взаимодействия между физическим и цифровым миром, накладывая цифровые элементы на физическую среду или создавая виртуальные среды с физическими элементами. Ключевым компонентом для смешанной реальности являются системы 3D сканирования, которые захватывают форму и текстуру объектов или сцен в виде облаков точек. Для эффективного использования этих данных в приложениях смешанной реальности, их преобразуют в полигональные сетки, подходящие для рендеринга, анимации и взаимодействия. Предлагаемый метод включает применение ResNet блоков для извлечения признаков, использование автоэнкодера для получения латентных 3D форм, и определение геометрии стен и потолков с помощью ограничивающих рамок. Метод позволяет получить полную 3D модель сцены из облака точек с помощью глубокого обучения и геометрического анализа. В данной работе использовались два набора данных для обучения и экспериментов: ShapeNet и ScanNet. Эти наборы данных представляют собой большую и разнообразную коллекцию трехмерных объектов и отсканированных сцен с подробной аннотацией.

Ключевые слова

3D-реконструкция, облако точек, глубокое обучение, геометрический анализ.

Deep Learning-based Polygonal Mesh Reconstruction from Point Clouds for Mixed Reality Visualization

M.I. Sorokin¹, D.D. Zhdanov¹, A.D. Zhdanov¹

¹ ITMO University, Kronverksky Pr. 49, bldg. A, St. Petersburg, 197101, Russia

Abstract

Mixed reality systems create new forms of interaction between the physical and digital world by overlaying digital elements on the physical environment or creating virtual environments with physical elements. A key component for mixed reality are 3D scanning systems that capture the shape and texture of objects or scenes in the form of point clouds. For effective use of these data in mixed reality applications, they are transformed into polygonal meshes suitable for rendering, animation, and interaction. The proposed method includes the application of ResNet blocks for feature extraction, the use of an autoencoder to obtain latent 3D shapes, and the definition of wall and ceiling geometry using bounding boxes. The method allows obtaining a complete 3D model of the scene from a point cloud using deep learning and geometric analysis. In this work, two datasets were used for training and experiments: ShapeNet and ScanNet. These datasets represent a large and diverse collection of three-dimensional objects and scanned scenes with detailed annotation.

Keywords

3D reconstruction, point cloud, deep learning, geometric analysis.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: vergotten@gmail.com (М.И. Сорокин); ddzhdanov@mail.ru (Д.Д. Жданов); andrew.gtx@gmail.com (А.Д. Жданов)
ORCID: 0000-0001-9093-1690 (М.И. Сорокин); 0000-0001-7346-8155 (Д.Д. Жданов); 0000-0002-2569-1982 (А.Д. Жданов)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

Восстановление 3D сцены из облака точек является важной и сложной задачей компьютерного зрения, которая имеет множество приложений в различных областях, таких как робототехника, дополненная и виртуальная реальность, оцифровка культурного наследия и т.д. Однако, облака точек, полученные с помощью лидара или другого датчика глубины, часто содержат шум, пропуски и выбросы, которые затрудняют точное и полное восстановление 3D сцены. Кроме того, облака точек имеют большой объем данных и не имеют явной структуры или топологии, что требует эффективных и устойчивых методов обработки. В этой работе разработан метод восстановления 3D сцены из облака точек, который позволяет получить полную и детальную 3D модель сцены в виде полигональных сеток. Полигональные сетки - это популярный формат для представления трехмерных объектов, который состоит из вершин, ребер и граней, формирующих сетчатую структуру поверхности объекта. Полигональные сетки удобны для дальнейшей обработки и визуализации, так как они содержат информацию о топологии и связности поверхности объекта.

Данный метод объединяет преимущества глубокого обучения и геометрического анализа для получения качественных полигональных сеток из облака точек. Глубокое обучение - это мощный инструмент для извлечения признаков и представлений из сложных данных, таких как облака точек. Геометрический анализ - это способ изучения формы и структуры объектов на основе математических методов. Сочетание этих двух подходов позволяет не только учитывать локальные и глобальные свойства облака точек, но и эффективно генерировать полигональные сетки. Основная идея метода заключается в том, что облако точек предварительно разбивается на небольшие фрагменты с помощью нейронной сети, и каждый фрагмент реконструируется в виде полигональной сетки. Полученные полигональные сетки склеиваются в единую 3D модель сцены. Этот процесс повторяется для разных уровней детализации, чтобы получить многоуровневую 3D модель сцены. Метод обладает хорошей обобщающей способностью и может адаптироваться к разным типам сцен и объектов.

2. Анализ текущего состояния

Для восстановления 3D формы из облака точек существует множество методов, которые можно разделить на две группы: методы на основе геометрии и методы на основе обучения. Методы на основе геометрии используют различные алгоритмы для построения триангуляций, выпуклых оболочек, альфа-форм или поверхностей поинтсет из облака точек. Примерами таких методов являются:

- Триангуляция Делоне [1] — это триангуляция для заданного множества точек на плоскости, при которой для любого треугольника все точки за исключением его вершин лежат вне окружности, описанной вокруг треугольника. Триангуляция Делоне максимизирует минимальный угол среди всех углов всех построенных треугольников.
- Построение выпуклой оболочки [2] — это задача построения минимального выпуклого множества, которое содержит все заданные точки. Выпуклая оболочка может быть построена с помощью различных алгоритмов, таких как алгоритм Грэхема, алгоритм Джарвиса, алгоритм «разделяй и властвуй» или алгоритм «заворачивания подарка». Выпуклая оболочка имеет множество приложений в вычислительной геометрии, таких как поиск ближайшего соседа, определение диаметра множества точек, построение триангуляции Делоне и диаграммы Вороного.
- Поиск альфа-форм [3] — это метод построения обобщенной выпуклой оболочки, которая может быть более гладкой и ближе к форме объекта, чем триангуляция Делоне. Альфа-форма определяется с помощью параметра альфа, который задает радиус сферы, используемой для удаления лишних ребер из триангуляции Делоне.
- Построение поверхности поинтсет [4] — это метод восстановления гладкой поверхности из облака точек, основанный на проекции точек на локальные касательные

пространства. Построение поверхности поинтсет учитывает нормали точек и может обрабатывать зашумленные и неполные данные.

Методы на основе обучения основаны на использовании нейронных сетей для извлечения признаков из облака точек и восстановления 3D формы в разных форматах: облако точек, треугольная сетка, воксели или SDF. Примерами таких методов являются:

- PointNet [5] — это первая нейронная сеть, которая работает непосредственно с облаками точек и учитывает перестановочную инвариантность точек во входных данных. PointNet предоставляет единый архитектуру для задач классификации объектов, сегментации частей и семантического анализа сцен.
- AtlasNet [6] — это метод, который генерирует треугольные сетки из облаков точек, используя набор параметрических поверхностей (атласов), которые отображаются на 3D форму объекта. AtlasNet учится выравнивать атласы с геометрией объекта и минимизировать разрывы между ними.
- PCN [7] — это метод, который восстанавливает полную 3D форму объекта из неполного облака точек, используя три последовательных модуля: дополнение, усиление и уточнение. PCN учится заполнять пропущенные части объекта и увеличивать плотность точек.
- 3D-LMNet [8] — это метод, который кодирует облако точек в латентный вектор и декодирует его в SDF. 3D-LMNet использует графовую сверточную сеть для извлечения признаков из облака точек и рекуррентную нейронную сеть для генерации SDF.
- 3D-PSRNet [9] — это метод, который восстанавливает 3D форму объекта из одного RGB изображения, используя двухступенчатый подход. Сначала он генерирует грубое облако точек из изображения, а затем уточняет его с помощью PointNet++.
- CAPNet [10] — это метод, который восстанавливает 3D форму объекта из одного RGB изображения, используя каскадный подход. Он состоит из трех модулей: детектор ключевых точек, генератор кривых и генератор поверхностей. CAPNet учится предсказывать ключевые точки объекта на изображении, соединять их кривыми и заполнять поверхности между кривыми.

Однако эти методы требуют больших объемов обучающих данных и вычислительных ресурсов и не могут гарантировать гладкость и замкнутость поверхности. В отличие от существующих методов, представленный в данной работе метод сочетает в себе преимущества геометрических методов и методов, основанных на машинном обучении. Предлагаемый метод использует глубокое машинное обучение для извлечения признаков и получения латентных 3D форм, а также геометрический анализ для удаления шума, восстановления отсутствующих точек и извлечения схемы расположения объектов (планировки) сцены.

Метод, описанный в данной работе, имеет ряд преимуществ по сравнению с другими работами по восстановлению 3D сцены из облака точек. Во-первых, он способен извлекать планировку сцены из облака точек, определяя положение и форму стен и потолков с помощью ограничивающих рамок. Это позволяет более точно оценить размеры и расположение объектов в пространстве. Во-вторых, он способен генерировать полигональные сетки высокого качества из латентных 3D форм, используя метод генерации облаков точек на основе FoldingNet [11]. Полигональные сетки лучше подходят для задач рендеринга, анимации и взаимодействия с виртуальными объектами в приложениях смешанной реальности. Однако, предлагаемый метод имеет ряд ограничений. Во-первых, он требует большого количества данных для обучения автоэнкодера и других компонентов сети. Во-вторых, он работает только с теми категориями объектов, которые использовались для обучения автоэнкодера. Это значит, что он не может восстанавливать объекты других категорий или объектов с сильно измененной формой.

3. Метод восстановления 3D сцены из облака точек

Предлагаемый метод состоит из таких этапов, как: извлечение признаков, получение латентных 3D форм, удаление шума и восстановление отсутствующих точек, извлечение

планировки сцены, генерация 3D формы объекта по латентным векторам, и построение 3D модели сцены:

- Извлечение признаков: на этом этапе из облака точек извлекаются локальные и глобальные признаки с помощью ResNet блоков [12], которые являются сверточными нейронными сетями с остаточными соединениями. Для этого сначала производится случайный отбор точек из облака до фиксированного размера N с помощью алгоритма farthest point sampling (FPS) [13]. Затем применяются ResNet блоки к каждой точке и ее k ближайшим соседям для получения локальных признаков. Локальные признаки характеризуют форму и текстуру поверхности в небольшой окрестности каждой точки. Для получения глобальных признаков применяются ResNet блоки ко всему облаку точек и используется операция max-pooling для агрегации информации. Глобальные признаки описывают структуру и контекст всего облака точек. Локальные и глобальные признаки конкатенируются для получения окончательного признакового представления облака точек размерности $N \times D$, где D - размерность признакового пространства. Размерность признакового пространства для облака точек - это количество признаков, которые используются для описания каждой точки в облаке.
- Получение латентных 3D форм: на этом этапе из признаков облака точек получаются латентные 3D формы объектов в сцене. Для этого используется автоэнкодер, который состоит из двух частей: кодировщика и декодировщика. Кодировщик принимает на вход признаки облака точек и выдает латентный вектор для каждого объекта. Для этого используется метод кластеризации на основе плотности DBSCAN [14], который разбивает облако точек на группы точек, принадлежащих одному объекту. Затем применяются ResNet блоки к каждому кластеру и используется операция max-pooling для получения латентного вектора размерности E , где E - размерность латентного пространства. Латентный вектор содержит компактную и низкоразмерную информацию о 3D форме объекта. Декодировщик принимает на вход латентный вектор и восстанавливает 3D форму объекта в виде облака точек или другого формата. Для этого используется метод генерации облаков точек на основе FoldingNet [11], который преобразует латентный вектор в матрицу координат точек размерности $M \times 3$, где M - количество точек в облаке. Метод FoldingNet основан на идее складывания 2D сетки до 3D формы с помощью нескольких полносвязных слоев.
- Удаление шума и восстановление отсутствующих точек: на этом этапе качество латентных 3D форм повышается с помощью дополнительных веток автоэнкодера, специализирующихся на разных аспектах: одна фильтрует или регуляризует облака точек для удаления шума; другая интерполирует или генерирует точки для заполнения пропусков. Для фильтрации шума применяется метод Topology-Aware Denoising (TAD) [15], который анализирует и устраняет выбросы на основе топологических свойств поверхности. Для восстановления отсутствующих точек используется метод Point Completion Network (PCN) [7], который дополняет облака точек до желаемого размера с помощью генеративной нейронной сети.
- Извлечение планировки сцены: на этом этапе для облака точек, снятого в помещении, определяется положение и форма стен и потолков с помощью ограничивающих рамок. Для этого применяется метод распознавания плоскостей или кластеризации точек по геометрическим свойствам. Используется нейронная сеть PlaneRCNN [16], которая одновременно детектирует, сегментирует и нормализует плоскости из облаков точек. Метод PlaneRCNN дает ограничивающие рамки для каждой плоскости в сцене, а также маски и нормали для каждой точки, относящейся к плоскости. Планировка сцены помогает более точно оценить размеры и расположение объектов в пространстве.
- Получение готовой 3D формы объекта по латентным векторам: на этом этапе из латентных векторов получаются готовые 3D формы объектов в виде полигональных сеток. Для этого используется нейронная сеть, которая генерирует полигональные сетки из латентных векторов. Нейронная сеть принимает на вход латентный вектор и выдает набор плоскостей, которые образуют полигональную сетку. Каждая плоскость

представляет собой квадрат, который деформируется с помощью нескольких сверточных слоев. Нейронная сеть может генерировать разнообразные и сложные 3D формы из латентных векторов.

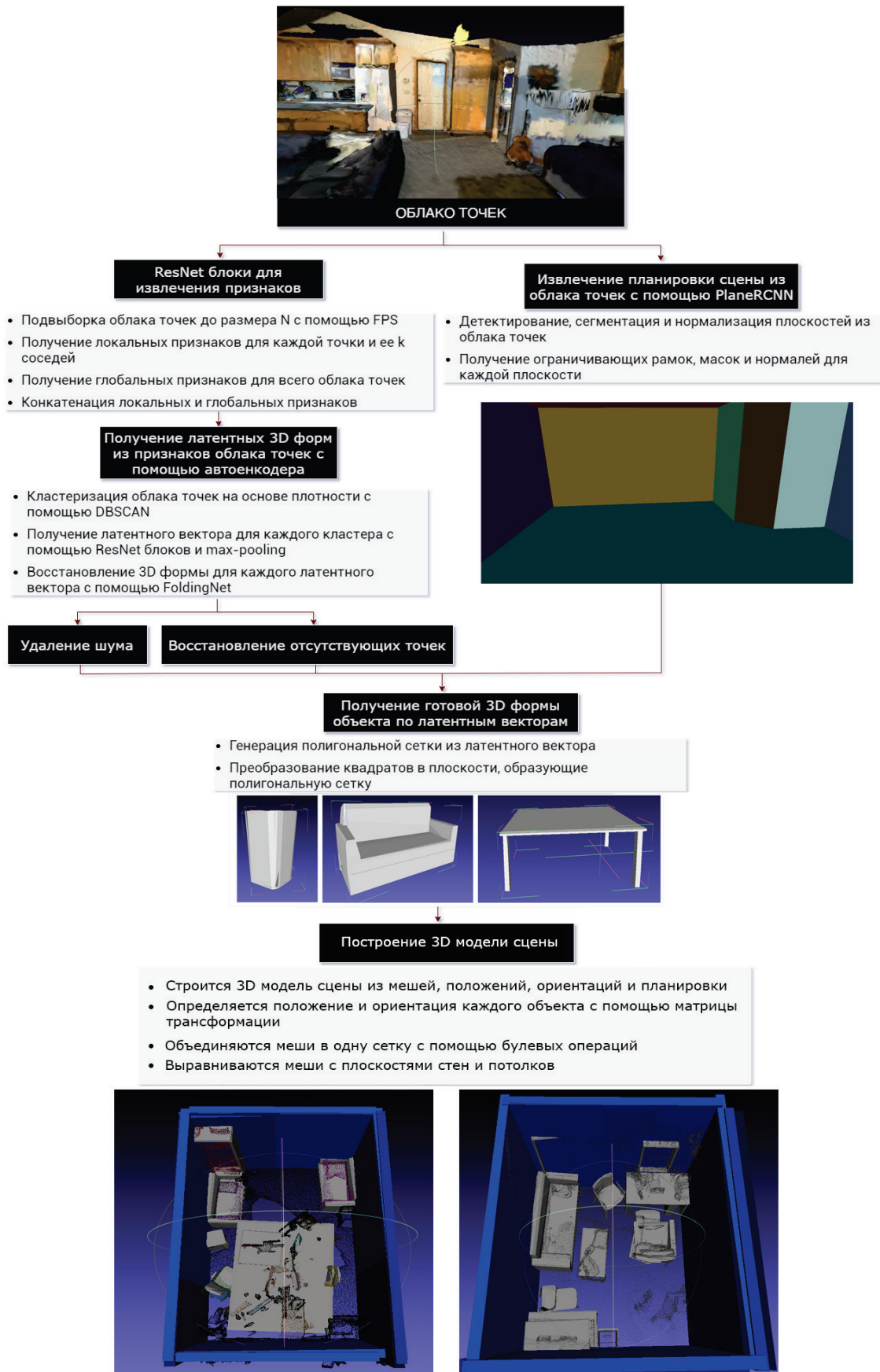
- Построение 3D модели сцены: на этом этапе из предсказанных положений треугольных сеток, ориентации в пространстве, планировки сцены и готовых 3D форм объектов строится 3D модель сцены в виде полигональной сетки. Для этого используется метод слияния треугольных сеток или объединения плоскостей. Метод работает следующим образом: для каждого объекта в сцене определяется его положение и ориентация в пространстве с помощью матрицы трансформации, которая применяется к его 3D форме. Затем все треугольные сетки объединяются в одну полигональную сетку с помощью булевых операций, таких как объединение, пересечение и разность. Метод также учитывает планировку сцены и корректирует положение и форму треугольных сеток так, чтобы они соответствовали плоскостям стен и потолков. Результатом является 3D модель сцены, которая содержит все предсказанные объекты и их формы, расположение и ориентацию в пространстве. На рисунке 1 показана общая схема предложенного метода.

4. Результаты и анализ эффективности предложенного метода

Для оценки качества полученных 3D форм объектов в виде полигональных сеток, в этой работе используются следующие метрики: AP_mesh, Precision_mesh и Recall_mesh. Эти метрики позволяют измерить, насколько хорошо предложенный метод детектирует и восстанавливает объекты из облака точек, используя латентные векторы.

AP_mesh - это средняя точность для полигональных сеток, которая рассчитывается как среднее арифметическое точностей для разных порогов пересечения по объединению (IoU) между предсказанными и истинными мешами. Precision_mesh - это доля правильно детектированных объектов среди всех предсказанных, которая рассчитывается как отношение количества истинно положительных (TP) к сумме истинно положительных и ложноположительных (FP) мешей. Recall_mesh - это доля правильно детектированных объектов среди всех реальных, которая рассчитывается как отношение количества истинно положительных (TP) к сумме истинно положительных и ложноотрицательных (FN) мешей. Эти метрики позволяют измерить, насколько хорошо метод детектирует и восстанавливает объекты из облака точек, используя разные пороги пересечения по объединению (IoU) между предсказанными и истинными мешами. Эксперименты проводятся на двух наборах данных: ScanNet и ShapeNetv2. ScanNet - это набор данных RGB-D видео, содержащий 2,5 миллиона кадров в более чем 1500 сканах, аннотированных 3D положениями камеры, поверхностными реконструкциями и семантическими сегментациями на уровне экземпляров. ShapeNetv2 - это набор данных 3D CAD-моделей, содержащий 57 категорий объектов. Для тестирования была использована выборка, состоящая из 8 категорий объектов: стол, стул, книжная полка, диван, мусорное ведро, шкаф, дисплей и ванна (таблица 1). Эта выборка была получена с помощью метода TSDF Fusion [17], который реконструирует 3D поверхность объектов из набора RGB-D изображений. Метод TSDF Fusion использует Truncated Signed Distance Function (TSDF), которая представляет собой скалярное поле, хранящее расстояние от каждой точки в пространстве до ближайшей поверхности. Метод объединяет несколько TSDF из разных точек зрения в один общий TSDF, который затем преобразуется в полигональную сетку с помощью алгоритма Marching Cubes. Метод TSDF Fusion позволяет получать гладкие и герметичные сетки из зашумленных и неполных данных. На рисунке 2 продемонстрированы облако точек, эталонная сегментация на уровне экземпляров, и полученные 3D модели для сцены из набора ScanNet.

Рисунок 3 демонстрирует сегментированное облако точек и восстановленные полигональные сетки объектов, а также восстановление планировки сцены с помощью ограничивающих прямоугольников.



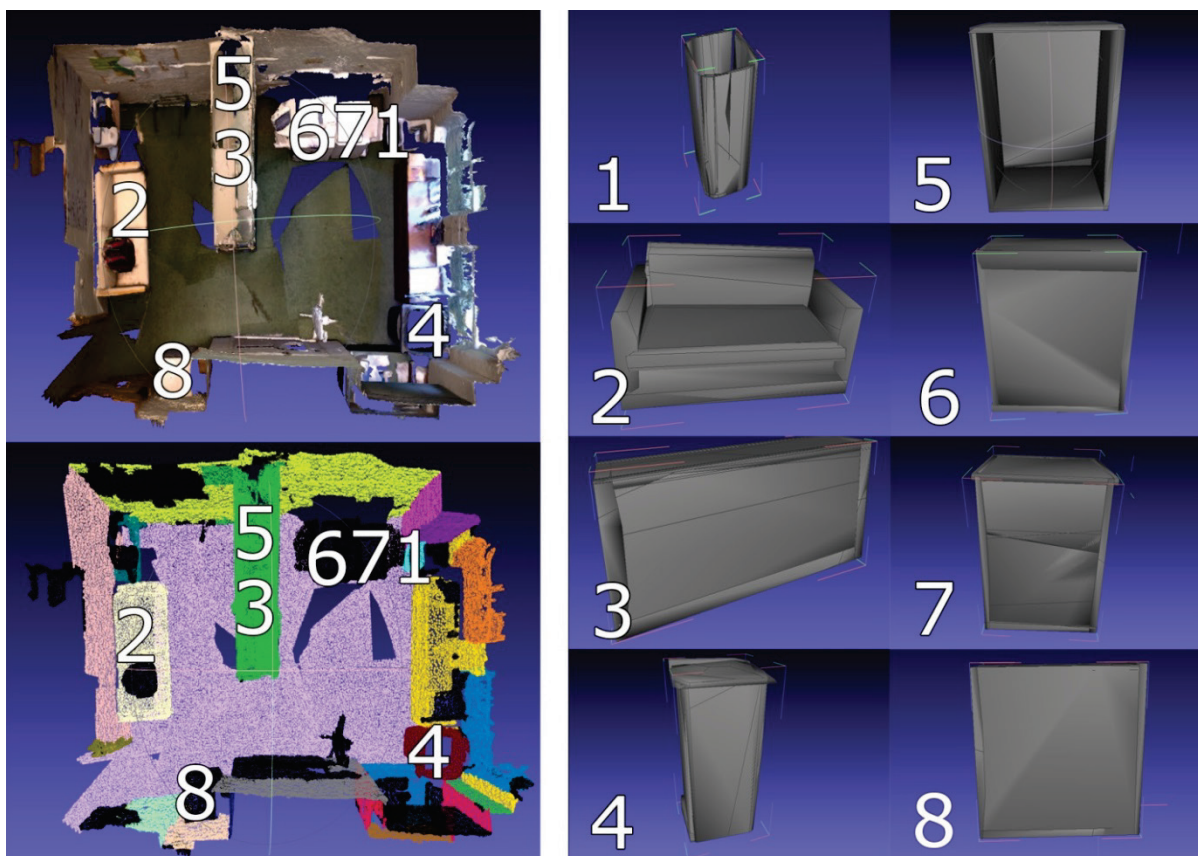


Рисунок 2 – Визуализация облака точек, эталонной сегментации и полученных 3D моделей для сцены из набора ScanNet

Таблица 1 – Метрики качества восстановления 3D формы для разных типов объектов при пороге IoU 0.1

Тип объекта	AP_mesh (%)	Precision_mesh (%)	Recall_mesh (%)
стол	37.455	44.518	55.372
стул	80.977	57.855	92.924
книжная полка	38.209	33.333	54.067
диван	44.099	41.727	55.238
мусорное ведро	62.604	47.619	84.112
шкаф	23.892	23.256	42.636
дисплей	75.123	61.786	87.500
ванна	50.928	55.455	52.586
Среднее	51.661	45.681	65.554

Для оценки и сравнения с другими методами используется mAP (средняя точность) в контексте 3D-реконструкции объектов. Этот показатель позволяет оценить качество восстановленных сеток путем сравнения их с истинными сетками с использованием метрики mesh IoU (Intersection over Union). При установке порога mesh IoU на 0,5, восстановленная сетка должна иметь перекрытие не менее 50% с истинной сеткой, чтобы считаться правильной реконструкцией. Это дает возможность оценить, насколько точно восстановленная сетка соответствует истинной форме объекта.



Рисунок 3 – Создание полигональных сеток объектов из облаков точек

Таблица 2 – Сравнение предложенного метода с аналогичными работами. Оценки mAP (средняя точность) измеряются с порогом mesh IoU (Intersection over Union) на уровне 0.5

	дисплей	ванна	мусорное ведро	диван	стул	стол	шкаф	книжная полка	mAp
RevealNet [18]	13.16	13.64	18.19	24.79	15.87	11.28	8.60	10.60	14.52
RfD-Net [19]	26.67	27.57	23.34	15.71	12.23	1.92	14.48	13.39	16.90
Предложенный метод	28.32	26.29	25.54	28.37	10.28	15.69	20.24	14.65	21.17

5. Заключение

В данной работе был предложен метод восстановления 3D сцен из облаков точек с помощью глубокого обучения и кластеризации. Метод состоит из нескольких этапов: восстановление 3D форм объектов, удаление шума и восстановление отсутствующих точек, извлечение планировки сцены и построение полигональной сетки. Метод был протестирован на различных сценах в помещении и показал достаточно корректные и эффективные результаты. Метод может быть полезен для различных приложений, связанных с 3D моделированием, анализом и визуализацией, а также для восстановления 3D сцен в системах смешанной реальности.

В качестве дальнейших направлений исследования планируется улучшить метод в следующих аспектах: обучение автоэнкодера для работы с большим количеством категорий, расширение для работы с динамическими сценами, использование других форматов представления 3D форм, таких как воксели, исследование возможности использования нескольких источников данных, таких как изображения или видео, для улучшения качества восстановления 3D сцен. Средняя точность восстановления 3D сцен при пороге IoU 0.1 составила около 50%, что является приемлемым результатом для данной задачи, учитывая сложность и разнообразие сцен, зашумленность и неполноту данных, а также ограниченность вычислительных ресурсов.

6. Финансирование

Работа была выполнена при финансовой поддержке Российского Научного Фонда, грант № 22-11-00145.

7. Список источников

- [1] Liu, N., Lin, B., Lv, G., Zhu, A. X., & Zhou, L. A Delaunay triangulation algorithm based on dual-spatial data organization. // PFG–Journal of Photogrammetry, Remote Sensing and Geoinformation Science. 2019. № 1-2. (pp. 19-31), doi:10.1007/s41064-019-00067-y
- [2] Алгоритмы вычислительной геометрии. Выпуклые оболочки: простые алгоритмы / С.А. Ивановский, А.С. Преображенский, С.К. Симончик // Компьютерные инструменты в образовании. 2007. № 6. (С.3-16).
- [3] Attali D., Lieutier A., Salinas L. Vietoris-Rips complexes also provide topologically correct reconstructions of sampled shapes // Proceedings of the twenty-seventh annual symposium on Computational geometry. ACM. 2011. (pp. 491-500), doi: 10.1145/1998196.1998276.
- [4] Guennebaud G., Gross M. Algebraic point set surfaces // ACM Transactions on Graphics (TOG). 2007. Vol. 26, No. 3. Article 23, doi: 10.1145/1275808.1276406.
- [5] Qi C.R., Su H., Mo K., Guibas L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation // Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE. 2017. (pp. 652–660).
- [6] Groueix T., Fisher M., Kim V.G., Russell B.C., Aubry M. Atlasnet: A papier-mâché approach to learning 3d surface generation // Proceedings of the IEEE conference on computer vision and pattern recognition. IEEE, 2018. (pp. 216–224).
- [7] Yuan W., Khot T., Held D., Mertz C., Hebert M. Pcn: Point completion network // 2018 International Conference on 3D Vision (3DV). IEEE, 2018, doi:10.1109/3DV.2018.00088
- [8] Sharma G., Goyal R., Liu D., Kalogerakis E., Maji S. Csgnet: Neural shape parser for constructive solid geometry // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018. (pp. 5515–5523).
- [9] Sharma G., Kalogerakis E., Maji S. 3d-psrnet: Part segmented 3d point cloud reconstruction from a single image // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. IEEE, 2019. (pp. 9620–9629).

- [10] Sharma G., Kalogerakis E., Maji S. Capnet: Continuous approximation projection for 3d point cloud reconstruction using 2d supervision // European Conference on Computer Vision. Springer, Cham, 2020. (pp. 24–40), doi:10.48550/arXiv.1811.11731
- [11] Yang Y., Feng C., Shen Y., Tian D. FoldingNet: Point Cloud Auto-encoder via Deep Grid Deformation // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2018. (pp. 206-215).
- [12] He K., Zhang X., Ren S., Sun J. Identity mappings in deep residual networks // European conference on computer vision. Springer, Cham, 2016. (pp. 630-645), doi:10.1007/978-3-319-46493-0_38
- [13] Mo K., Guerrero P., Yi L., Su H., Wonka P., Mitra N.J., Guibas L.J. StructureNet: hierarchical graph networks for 3D shape generation // ACM Transactions on Graphics (TOG). 2019. Vol. 38, No. 5. Article 169, doi:10.48550/arXiv.1908.00575
- [14] Achlioptas P., Diamanti O., Mitliagkas I., Guibas L.J. Learning representations and generative models for 3D point clouds // International Conference on Machine Learning. PMLR, 2018. (pp. 40-49).
- [15] Sun G., Chu C., Mei J., Bi W., Su Z. Topology-Aware Denoising for Real-world Noisy Point Clouds with Complex Structures // Computer-Aided Design. 2022. Vol. 149. Article 103275, doi:10.1016/j.cad.2022.103275
- [16] Liu C., Kim K., Gu J., Furukawa Y., Kautz J. PlaneRCNN: 3D Plane Detection and Reconstruction from a Single Image // IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2019. (pp. 4450-4459).
- [17] Stutz D., Geiger A. Learning 3D Shape Completion under Weak Supervision // CoRR, abs/1805.07290, 2018, doi:10.1007/s11263-018-1126-y.