

ПРЕОБРАЗОВАНИЕ ДАННЫХ ОБЛАКА ТОЧЕК В 3D-МОДЕЛИ  
С ПОМОЩЬЮ POINTNET++ И TRANSFORMER© 2024 г. М. И. Сорокин<sup>a,\*</sup>, Д. Д. Жданов<sup>a,\*\*</sup>, А. Д. Жданов<sup>a,\*\*\*</sup><sup>a</sup> Университет ИТМО

197101 Санкт-Петербург, Кронверкский проспект, д. 49, литер А, Россия

\*E-mail: [vergotten@gmail.com](mailto:vergotten@gmail.com)\*\*E-mail: [ddzhdanov@mail.ru](mailto:ddzhdanov@mail.ru)\*\*\*E-mail: [andrew.gtx@gmail.com](mailto:andrew.gtx@gmail.com)

Поступила в редакцию: 19.01.2024 г.

После доработки: 19.01.2024 г.

Принята к публикации: 24.01.2024 г.

В данной работе представлен подход к восстановлению 3D-моделей из данных облака точек, основанный на использовании современных архитектур нейронных сетей. Основу метода составляют PointNet++ и Transformer. PointNet++ играет центральную роль, обеспечивая эффективное извлечение признаков и кодирование сложных геометрий 3D-сцен. Это достигается благодаря рекурсивному применению PointNet++ к вложенным разбиениям входного набора точек в метрическом пространстве. Выпуклая декомпозиция, являющаяся важным этапом в подходе, позволяет преобразовать сложные трехмерные объекты в набор более простых выпуклых форм. Это упрощает обработку данных и делает процесс восстановления более управляемым. Transformer затем обучает модель на этих признаках, что позволяет генерировать высококачественные реконструкции. Важно отметить, что Transformer используется исключительно для определения положения стен и границ объектов. Это сочетание технологий позволяет достичь высокой точности при восстановлении 3D-моделей. Основная идея метода заключается в сегментации облака точек на малые фрагменты, которые затем восстанавливаются как полигональные сетки. Для восстановления отсутствующих точек в данных облака точек применяется метод, основанный на алгоритме L1-Median и локальных признаках облака точек. Этот подход может адаптироваться к различным геометрическим структурам и исправлять ошибки топологического соединения. Предложенный метод был сравнен с несколькими современными подходами и показал свой потенциал в различных областях, включая архитектуру, инжиниринг, цифровизацию культурного наследия и системы дополненной и смешанной реальности. Это подчеркивает его широкую применимость и значительный потенциал для дальнейшего развития и применения в различных областях.

**Ключевые слова:** облако точек, трехмерные модели, нейронные сети, PointNet++, трансформер, выпуклое разложение

**DOI:** 10.31857/S0132347424030044, **EDN:** QARZUT

## 1. ВВЕДЕНИЕ

В области компьютерного зрения преобразование облака точек в трехмерную сцену представляет собой сложную проблему. Это охватывает широкий спектр областей, включая робототехнику, дополненную и виртуальную реальность, а также оцифровку культурного наследия. Облака точек, которые обычно получают с помощью лидара или других датчиков глубины, часто содержат шум, пробелы и выбросы, что затрудняет процесс точной и полной реконструкции трехмерной сцены. Кроме того, облака точек требуют большого объема данных и не имеют внутренней структуры или топологии, что требует надежных и эффективных методов обработки.

В рамках данной работы разрабатывается метод реконструкции трехмерной сцены из облака точек, который позволяет создать детализированную и полную трехмерную модель сцены в виде полигональных сеток. Эти сетки, состоящие из вершин, ребер и граней, образующих сетчатую структуру на поверхности объекта, являются распространенным форматом представления трехмерных объектов. Они идеально подходят для дальнейшей обработки, поскольку предоставляют информацию о топологии и связности поверхности объекта.

Метод объединяет преимущества глубокого обучения и геометрического анализа для создания высококачественных полигональных сеток из облаков точек. Глубокое обучение отлично

подходит для извлечения функций и представлений из сложных данных, таких как облака точек, в то время как геометрический анализ дает математический подход к изучению формы и структуры объектов. Объединив эти два подхода, можно учитывать как локальные, так и глобальные свойства облака точек, одновременно эффективно создавая полигональные сетки.

Основная концепция метода заключается в первоначальном сегментировании облака точек на небольшие фрагменты с помощью нейронной сети. Затем каждый фрагмент реконструируется как полигональная сетка. Эти отдельные сетки впоследствии объединяются в единую трехмерную модель сцены. Эта процедура повторяется на различных уровнях детализации для создания многоуровневой трехмерной модели сцены.

Для улучшения процесса метод включает в себя использование герметичных сеток и выпуклого разложения. Герметичные сетки, представляющие собой многоугольные сетки без каких-либо отверстий, зазоров или свободных краев, идеально подходят для представления твердых объектов в трехмерном пространстве, поскольку они гарантируют, что каждое ребро принадлежит ровно двум граням. Это имеет решающее значение для сохранения целостности трехмерной модели.

Выпуклая декомпозиция — еще один ключевой компонент подхода, который включает в себя разбиение сложной трехмерной формы на набор более простых выпуклых фигур. С такими формами проще работать в вычислительном отношении и их можно обрабатывать более эффективно. В данном методе каждый фрагмент облака точек раскладывается на выпуклые формы, а затем реконструируется в виде полигональной сетки.

Используя герметичные полигональные сетки и выпуклую декомпозицию, метод может обрабатывать сложные трехмерные структуры и обеспечивать точность и полноту реконструированной трехмерной сцены. Это делает подход надежным и универсальным, демонстрируя сильные возможности обобщения и адаптивности к различным типам сцен и объектов.

## 2. АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ

Процесс создания трехмерных полигональных сетей из облаков точек является сложной задачей, которая требует детального анализа и извлечения объектов из данных облака точек. Этот процесс также включает в себя оптимизацию процесса обучения для генерации моделей объектов

и восстановление отсутствующих точек. Работа со сложной геометрией и неполными данными может представлять особые трудности, особенно в реальных условиях. Однако, благодаря применению передовых архитектур нейронных сетей и инновационных методов обучения, эти проблемы можно преодолеть, что позволяет создавать высококачественные 3D-реконструкции.

Есть множество подходов к реконструкции трехмерных форм из облаков точек, которые можно условно разделить на две категории: методы, основанные на геометрии, и методы, основанные на обучении. Геометрические методы используют разнообразные алгоритмы для создания триангуляций [1], выпуклых оболочек [2], альфа-форм [3] или поверхностей множества точек из облаков точек [4]. Эти методы обычно требуют значительных вычислительных ресурсов и могут быть сложными в реализации, но они обеспечивают высокую точность и могут быть очень эффективными в определенных условиях.

С другой стороны, методы, основанные на обучении, используют методы машинного обучения для анализа облаков точек и создания трехмерных моделей. Эти методы могут быть более гибкими и масштабируемыми, чем геометрические методы, и они особенно полезны при работе с большими наборами данных или сложными формами. Однако они также могут требовать значительного количества данных для обучения и в некоторых случаях быть менее точными. Поэтому выбор метода зависит от конкретной задачи и доступных ресурсов. В области воссоздания 3D-форм из облаков точек некоторые методы обучения показали впечатляющие результаты. Эти подходы применяют нейронные сети для выделения характеристик из облаков точек и восстановления трехмерных объектов в разнообразных форматах. Вот несколько передовых методов, которые используются для обработки и анализа облаков точек в трехмерном пространстве:

- PointNet [5] — нейронная сеть, способная напрямую работать с облаками точек. Она уникальна тем, что учитывает перестановочную инвариантность точек во входных данных. PointNet предлагает универсальную архитектуру, которая может быть применена для решения различных задач, включая классификацию объектов, сегментацию частей и семантический анализ сцен;

- AtlasNet [6] — представляет собой передовой метод, который преобразует облака точек в треугольные сетки с помощью коллекции параметрических поверхностей, называемых атласами. Эти

атласы затем проецируются на 3D-структуру объекта. AtlasNet обучается синхронизировать атласы с геометрией объекта и уменьшать промежутки между ними, что способствует созданию более точных и подробных моделей;

- PCN [7] – подход, который способен реконструировать полную 3D-структуру объекта из неполного облака точек. PCN применяет три последовательно действующих модуля: дополнение, усиление и уточнение. PCN обучается восполнять отсутствующие части объекта и увеличивать плотность точек, что позволяет создавать более полные и детализированные модели.

В контексте изучения семантики и формы объектов на основе разреженных облаков точек RfD-Net [8] представляет собой инновационный подход. Эта архитектура разделяет процесс реконструкции семантического экземпляра на две ключевые задачи: глобальную локализацию объекта и прогнозирование локальной формы. RfD-Net разработан для работы с неполными облаками точек 3D-сцены, целью является одновременное понимание структуры 3D-объектов и восстановление сеток экземпляров объектов. Он использует разреженность данных облаков точек, уделяя особое внимание прогнозированию форм с высокой степенью детализации. Одной из особенностей RfD-Net является его способность справляться с проблемой увеличения сложности модели, что часто встречается в существующих методах. Это достигается за счет использования генератора форм, поддерживающего прямое неявное обучение. В каждом сегменте объекта облака точек содержатся детали формы, которые помогают неявному обучению функций, позволяя реконструировать поверхности с высокой степенью дискретизации. Еще один подход, реализующий принцип “сквозной обработки” (end-to-end), который заслуживает особого внимания, – это RevealNet [9]. Эта архитектура нейронной сети применяет методы совместного обучения для изучения характеристик цвета и геометрии. Она специально разработана для работы с неполными сканами RGB-D-сцены в целях идентификации отдельных экземпляров объектов и определения их полной геометрической структуры. RevealNet обладает полносверточной 3D-структурой, что обеспечивает эффективное семантическое завершение экземпляра для 3D-сканирования больших помещений за один проход. В результате сканируемая сцена разбивается на отдельные и полные трехмерные объекты, включая скрытые и невидимые части объекта, что демонстрирует потенциал и гибкость при работе

с большими и сложными наборами данных. Эти методы представляют собой важные шаги в направлении более эффективного и точного анализа и восстановления трехмерных форм из облаков точек. Они открывают новые возможности для исследований и разработок в области компьютерного зрения и машинного обучения.

В данной работе представлен метод, который эффективно объединяет преимущества геометрических подходов и подходов, основанных на машинном обучении. Этот метод использует принципы глубокого обучения для выделения характеристик и восстановления скрытых трехмерных форм, в то время как геометрический анализ применяется для устранения шума, восстановления отсутствующих точек и извлечения структуры объекта из сцены. Одним из основных достоинств этого подхода, по сравнению с другими методами восстановления 3D-сцены из облаков точек, является его способность воспроизводить планировку сцены. Используя сеть, основанную на трансформере, адаптированную в соответствии с такими исследованиями, как PQ-Transformer [10] и Omni-PQ [11], можно определить положение и форму стен и потолков с помощью ограничивающих рамок. Это обеспечивает более точную оценку размера объекта и его расположения в пространстве. Более того, этот подход способен создавать полигональные сетки из скрытых 3D-форм, используя адаптированный BSP-Net [12]. Однако у этого подхода есть определенные ограничения. Для обучения автоэнкодера и других компонентов сети необходим значительный объем данных. Более того, этот подход совместим только с теми категориями объектов, которые были использованы при обучении. Следовательно, он не способен реконструировать объекты из других категорий или объекты с существенно измененной формой. В рамках применяемого метода используется алгоритм L1-Median [13] для выявления ключевых точек в структуре облака точек. Затем, с использованием этих ключевых точек, определяется главное направление и рассчитывается плотность локального облака точек. Этот процесс создает условия для дополнения недостающих точек, что способствует более полному отображению структуры облака точек.

### 3. ПОДХОД К РЕКОНСТРУКЦИИ 3D-СЦЕНЫ ИЗ ОБЛАКА ТОЧЕК

Предложенный метод представляет собой последовательность ключевых этапов. Начальный этап – это извлечение признаков, включающее

в себя обработку данных и выделение значимых характеристик для последующей сегментации. Следующий этап – формирование скрытых трехмерных форм из объектов. После этого происходит анализ и извлечение структуры сцены. Затем, на основе скрытых векторов, создаются трехмерные модели объектов. И наконец, все данные и модели объединяются для создания окончательной 3D-модели сцены. Этот процесс обеспечивает целостный подход к реконструкции 3D-сцены из облака точек.

В начальной фазе процесса извлечения признаков с использованием PointNet++ [14], усовершенствованной версии PointNet, происходит детектирование локальных и глобальных объектов из облака точек. PointNet++ строит иерархические отношения между точками, что позволяет лучше улавливать локальные структуры и обобщать сложные сцены. Этот процесс включает в себя несколько важных шагов. Сначала из облака точек случайным образом отбираются точки до фиксированного размера  $N$  с применением алгоритма выборки самой дальней точки (FPS). Затем каждая отобранная точка и ее  $k$  ближайших соседей подвергаются обработке для получения локальных объектов. Эти локальные характеристики отражают форму и текстуру поверхности в непосредственной близости от каждой точки. Для получения глобальных объектов проводится обработка всего набора точек, при этом для сбора информации используется операция агрегирования. Эти глобальные характеристики описывают

структуру и контекст всего набора. В конце локальные и глобальные объекты объединяются для формирования окончательного представления объекта в виде набора точек размером  $N \times D$ , где  $D$  – размерность пространства объектов. Таким образом, начальный этап извлечения признаков является критическим для дальнейшего анализа и сегментации данных набора точек.

После формирования окончательных представлений объектов в виде облака точек, следующим шагом является семантическая сегментация. В рамках этого процесса каждая точка в облаке точек классифицируется на основе извлеченных признаков. Эти сгруппированные точки проходят через ряд слоев для прогнозирования класса объекта для каждой точки. В процессе извлечения признаков и сегментации используется функция перекрестной энтропии. Эта функция потерь измеряет разницу между предсказаниями модели и истинными метками. Для оптимизации процесса используется оптимизатор Адама, который адаптивно корректирует скорость обучения для каждого параметра, основываясь на оценках первого и второго моментов градиентов.

Эксперименты осуществлялись на двух наборах данных: ScanNet и ShapeNetv2. ScanNet представляет собой набор RGB-D-данных, включающий в себя 2.5 млн кадров, полученных в результате более чем 1500 сканирований. Эти данные аннотированы позициями 3D-камер, реконструкциями поверхности и семантической сегментацией на уровне отдельных экземпляров.

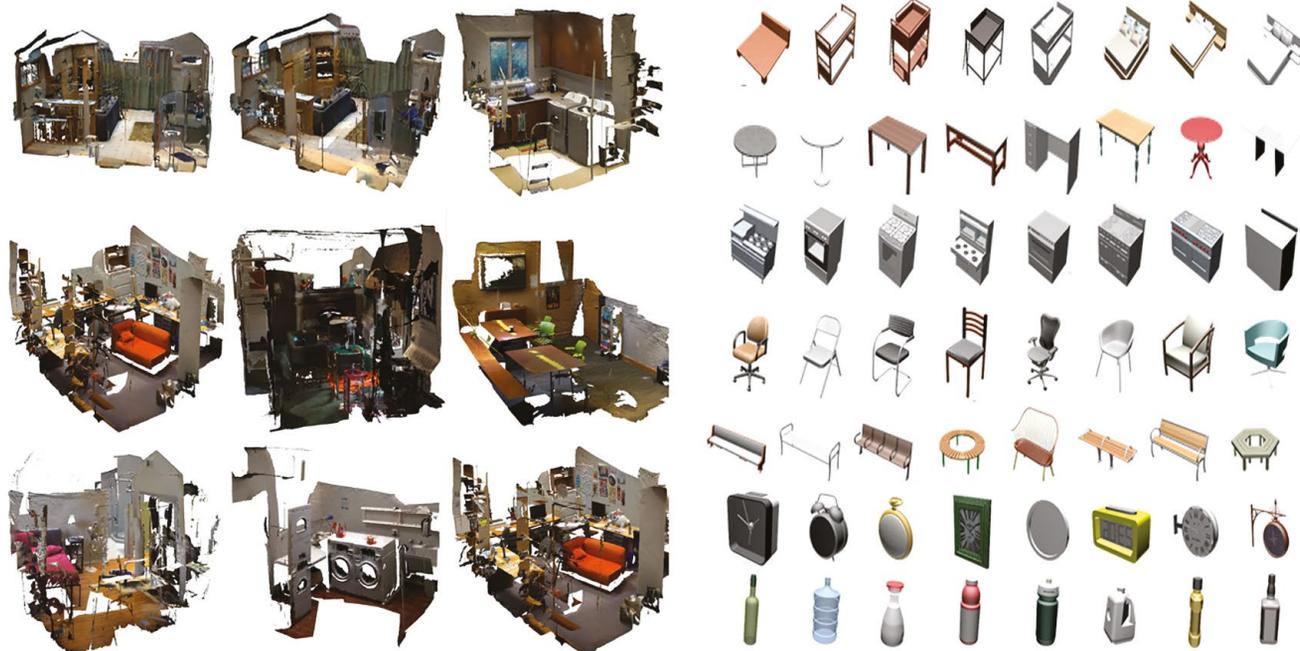


Рис. 1. Наборы данных ScanNet и ShapeNetv2.



Рис. 2. Результаты семантической сегментации на уровне классов и экземпляров сцен.

В свою очередь, ShapeNetv2 является набором данных 3D-моделей, включающим в себя 57 категорий объектов (рис. 1).

На рис. 2 представлены результаты семантической сегментации на уровне классов и сегментации на уровне экземпляров сцен.

Слева направо рисунок демонстрирует:

- исходное облако точек – первичные 3D-данные, подлежащие анализу;
- семантический прогноз на уровне классов – определение класса, к которому относится каждая точка в облаке;
- семантические истинные метки классов для каждой точки, используемые для обучения и проверки модели;
- семантический прогноз на уровне экземпляров – определение, к какому конкретному отдельному объекту принадлежит каждая точка;
- семантические истинные метки объекта для каждой точки на уровне экземпляров, используемые для обучения и проверки модели.

На этапе извлечения макета облака точек преобразуются в прямоугольные блоки с помощью сети трансформера. Модуль голосования генерирует прогнозы блоков. Затем эти прогнозы обрабатываются декодером трансформера для дальнейшего уточнения. После нескольких слов прямого распространения и подавления не-

максимальных значений (NMS) эти прогнозы становятся окончательными ограничивающими рамками, которые определяют положения и размеры стен и потолков в пространстве. На рис. 3 демонстрируются результаты извлечения макета. Прямоугольники, которые символизируют стены сцен, кодируются цветом. Зеленый цвет соответствует истинным положительным прогнозам, в то время как желтый цвет указывает на ложные положительные прогнозы.

Ключевым элементом подхода является выпуклое разложение на герметичных сетках, полученных из набора данных ShapeNet с помощью TSDF Fusion [15]. Сеть, обученная представлять 3D-форму через выпуклое разложение, использует принцип бинарного разделения пространства

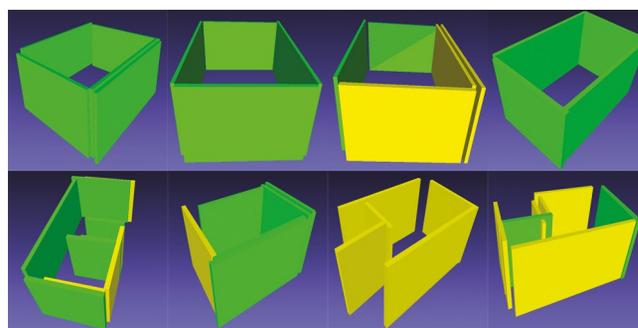


Рис. 3. Визуализация результатов извлечения макета.

(BSP), который заключается в рекурсивном разделении пространства для формирования выпуклых множеств. В процессе обучения сеть кодирует 3D-форму в скрытый вектор, который захватывает основные особенности входной формы. Декодер сети затем принимает этот скрытый вектор и декодирует его в набор параметров, которые определяют BSP-дерево. Это BSP-дерево используется для генерации набора выпуклых тел, которые аппроксимируют исходную 3D-форму. Сеть обучается восстанавливать форму, минимизируя функцию потерь между истинной формой и формой, восстановленной сетью. Потери восстановления вычисляются как среднеквадратичная ошибка между истинными и предсказанными значениями плотности. Это означает, что потери восстановления штрафуют модель за разницу между истинной формой и формой, восстановленной сетью. Градиенты функции потерь вычисляются и распространяются обратно через сеть. Затем параметры сети обновляются с использованием алгоритма оптимизации Adam для минимизации потерь. Результатом являются компактные полигональные сетки, которые идеально подходят для представления острых геометрических форм, гарантированно герметичны и могут быть легко параметризованы. На рис. 4 показаны предсказанные и истинные 3D-модели.



Рис. 4. Сравнение предсказанных (слева) и истинных (справа) 3D-моделей.

После определения формы, ориентации и расположения объектов на сцене эти данные собираются в полноценную 3D-модель. Модель, представляющая собой полигональную сетку, создается на основе предсказанных 3D-форм объектов и прямоугольников макета. Пример такой интегрированной 3D-модели сцены можно увидеть на рис. 5.

На рис. 6 представлены сегментированные облака точек и их восстановленные 3D-модели сцен.

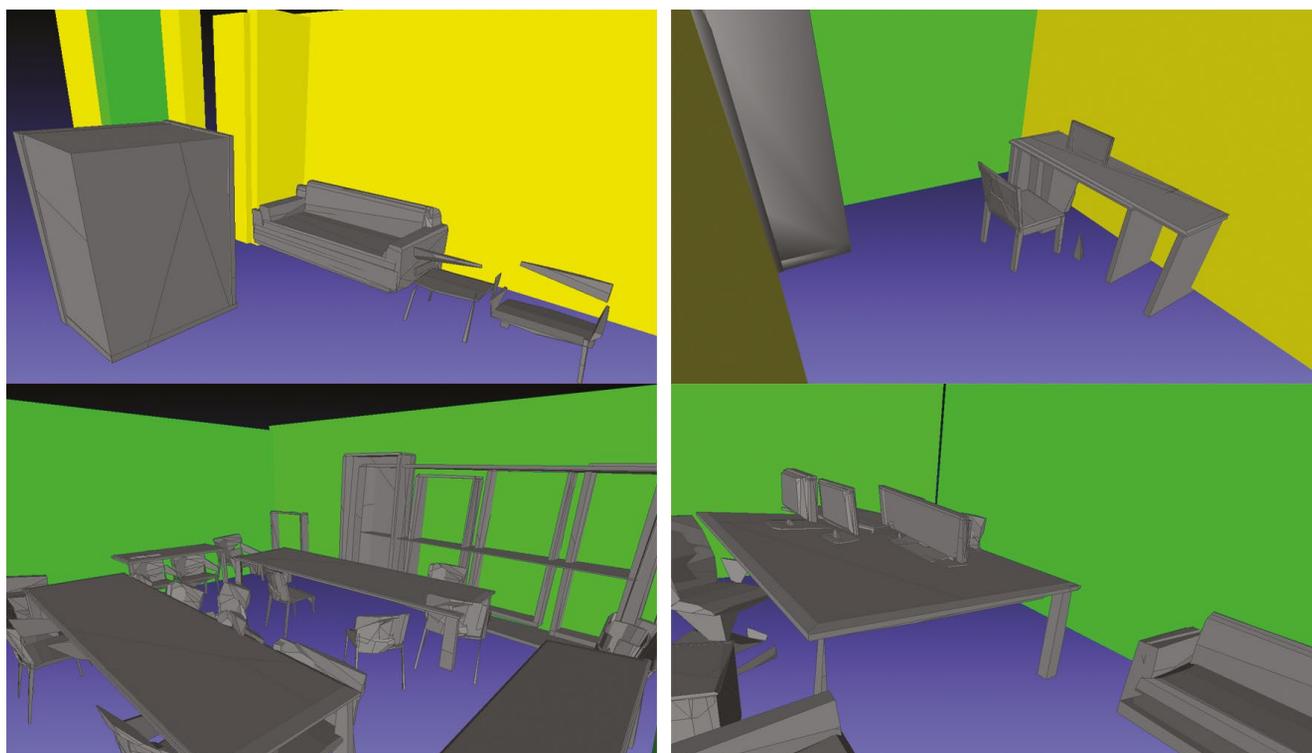


Рис. 5. Формирование 3D-модели сцены из предсказанных данных.



Рис. 6. Сегментированные облака точек и восстановленные 3D-модели сцен.

#### 4. ОЦЕНКА ЭФФЕКТИВНОСТИ ПРЕДЛОЖЕННОГО МЕТОДА

В данной работе для оценки качества получаемых 3D-форм объектов, представленных в виде полигональных сеток, применяются метрики *AP\_mesh*, *Precision\_mesh* и *Recall\_mesh*. Эти метрики позволяют оценить эффективность предложенного метода в обнаружении и восстановлении объектов из облаков точек с использованием скрытых векторов.

*AP\_mesh* представляет собой среднюю точность для полигональных сеток, которая вычисляется как среднее арифметическое точностей для различных порогов пересечения по объединению (IoU) между предсказанными и истинными сетками. *Precision\_mesh* определяет долю правильно обнаруженных объектов среди всех предсказанных, вычисляемую как отношение числа истинно положительных (TP) к сумме истинно положительных и ложно положительных (FP) сеток.

*Recall\_mesh*, в свою очередь, представляет собой долю правильно обнаруженных объектов среди всех реальных, вычисляемую как отношение числа истинно положительных (TP) к сумме истинно положительных и ложно отрицательных (FN) сеток. Эти метрики позволяют оценить, на-

сколько хорошо метод обнаруживает и восстанавливает объекты из облаков точек с использованием различных порогов пересечения по объединению (IoU) между предсказанными и истинными сетками.

В процессе тестирования использовалась выборка, включающая в себя восемь категорий объектов: стол, стул, книжная полка, диван, мусорное ведро, шкаф, дисплей и ванна (табл. 1). Для формирования этой выборки применялся метод *TSDF Fusion*, который позволяет восстанавливать 3D-поверхности объектов на основе набора *RGB-D*-изображений. Метод *TSDF Fusion* применяет специальное скалярное поле, которое хранит расстояние от каждой точки в пространстве до ближайшей поверхности. Этот метод объединяет несколько таких полей, полученных с разных точек зрения, в одно общее поле. Затем общее поле преобразуется в полигональную сетку с помощью алгоритма *Marching Cubes*. Благодаря методу *TSDF Fusion* возможно получение гладких и водонепроницаемых сеток из шумных и неполных данных.

Для оценки и сравнения с другими методами в контексте 3D-реконструкции объектов используется метрика *mAP* (средняя средняя точность).

**Таблица 1.** Метрики качества восстановления 3D-формы для различных 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

**Таблица 2.** Сравнение предложенного метода с аналогичными работами. Оценки mAP (средняя точность) измеряются при пороге IoU для сетки 0.5

	Дисплей	Ванна	Мусорное ведро	Диван	Стул	Стол	Шкаф	Книжная полка	mAp
RevealNet	13.16	13.64	18.19	24.79	15.87	11.28	8.60	10.60	14.52
RfD-Net	26.67	27.57	23.34	15.71	12.23	1.92	14.48	13.39	16.90
<b>Наш метод</b>	<b>28.32</b>	<b>26.29</b>	<b>25.54</b>	<b>28.37</b>	<b>10.28</b>	<b>15.69</b>	<b>20.24</b>	<b>14.65</b>	<b>21.17</b>

Эта метрика позволяет оценить качество восстановленных сеток путем их сравнения с эталонными сетками с использованием метрики IoU (Intersection over Union) для сеток. При установке порога IoU для сеток на уровне 0.5 восстановленная сетка должна перекрывать по крайней мере 50% эталонной сетки, чтобы считаться правильно восстановленной. Это позволяет оценить, насколько точно восстановленная сетка соответствует истинной форме объекта (табл. 2).

## 5. ЗАКЛЮЧЕНИЕ

В данной работе предлагается подход к созданию 3D-моделей на основе данных облака точек, который успешно решает задачи расположения элементов, позиционирования объекта и восстановления полигональных сеток. Данный метод применяет передовые архитектуры нейронных сетей, включая PointNet++ для извлечения признаков. Это позволяет модели улавливать сложные зависимости и отношения между различными частями данных, что в свою очередь помогает в создании более точных и детализированных 3D-моделей. PointNet++ является эффективной нейронной сетью, способной обрабатывать облака точек напрямую, захватывая сложные геометрические структуры, характерные для 3D-данных. Transformer, в свою очередь, представляет собой модель, которая использует механизмы самовнимания для уделения внимания различным частям входных данных при вычислении представления данных.

В этом контексте Transformer применяется для извлечения макета из данных. Ключевым элементом данного метода является выпуклое декомпозирование, т. е. процесс, в ходе которого сложная форма разбивается на более простые выпуклые части. Выпуклые формы обладают более простыми математическими свойствами, которые можно использовать для эффективных вычислений, что делает их более удобными для работы в различных приложениях, таких как компьютерная графика, физические симуляции, рендеринг и 3D-печать. Используя эти техники в совокупности, предложенный метод эффективно кодирует сложные геометрии 3D-сцен и генерирует реконструкции. Кроме того, он решает проблему отсутствующих точек, применяя специальные методы для восстановления этих точек. Предложенный метод продемонстрировал обещающие результаты и имеет потенциал для применения в различных областях, включая архитектуру, инженерию, цифровизацию культурного наследия, а также системы дополненной и смешанной реальности и другие области, где требуется воссоздание 3D-сцен.

## 6. ИСТОЧНИК ФИНАНСИРОВАНИЯ

Эта работа была поддержана Российским научным фондом, грант № 22-11-00145.

## СПИСОК ЛИТЕРАТУРЫ

1. Liu N., Lin B., Lv G., Zhu A.X., Zhou L. A Delaunay triangulation algorithm based on dual-spatial data orga-

- nization // PFG—Journal of Photogrammetry, Remote Sensing and Geoinformation Science. 2019. P. 19–31.
2. *Ivanovsky S.A., Preobrazhensky A.S., Simonchik S.K.* Algorithms for computational geometry. Convex hulls: simple algorithms // Computer tools in education. 2007. P. 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. P. 491–500.
  4. *Guennebaud G., Gross M.* Algebraic point set surfaces. ACM Transactions on Graphics (TOG). 2007. V. 26. No. 3. Article 23.
  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. P. 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. P. 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. <https://doi.org/10.1109/3DV.2018.00088>
  8. *Yinyu Nie, Ji Hou, Xiaoguang Han, Matthias Nießner.* RfD-Net: Point Scene Understanding by Semantic Instance Reconstruction // Computer Vision and Pattern Recognition. 2020.
  9. *Ji Hou, Angela Dai, Matthias Nießner.* RevealNet: Seeing Behind Objects in RGB-D Scans // Computer Vision and Pattern Recognition. 2019.
  10. *Xiaoxue Chen, Hao Zhao, Guyue Zhou, Ya-Qin Zhang.* PQ-Transformer: Jointly Parsing 3D Objects and Layouts from Point Clouds // Computer Vision and Pattern Recognition. 2021.
  11. *Huan-ang Gao, Beiwen Tian, Pengfei Li, Xiaoxue Chen, Hao Zhao, Guyue Zhou, Yurong Chen, Hongbin Zha.* From Semi-supervised to Omni-supervised Room Layout Estimation Using Point Clouds // Computer Vision and Pattern Recognition. 2023.
  12. *Zhiqin Chen, Andrea Tagliasacchi, Hao Zhang.* Learning Mesh Representations via Binary Space Partitioning Tree Networks // Computer Vision and Pattern Recognition. 2021.
  13. *Wei Cao, Jiayi Wu, Yufeng Shi, Dong Chen.* Restoration of Individual Tree Missing Point Cloud Based on Local Features of Point Cloud. Terrestrial and Mobile Mapping in Complex Indoor and Outdoor Environments. 2022.
  14. *Qi C.R., Yi L., Su H., Guibas L.J.* PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space // Proceedings of the Computer Vision and Pattern Recognition. 2017.
  15. *Zhaofeng Niu, Yuichiro Fujimoto, Masayuki Kanbara, Taishi Sawabe, Hirokazu Kato.* DFusion: Denoised TSDF Fusion of Multiple Depth Maps with Sensor Pose Noises // Computer Vision and Machine Learning for Intelligent Sensing Systems. 2022.

## CONVERSION OF POINT CLOUD DATA TO 3D MODELS USING POINTNET++ AND TRANSFORMER

© 2024 M. I. Sorokin<sup>a</sup>, D. D. Zhdanov<sup>a</sup>, A. D. Zhdanov<sup>a</sup>

<sup>a</sup>ITMO University

Kronverkskii pr. 49A, St. Petersburg, 197101 Russia

This work presents an approach to reconstructing 3D models from point cloud data, based on the use of modern neural network architectures. The basis of the method is PointNet++ and Transformer. PointNet++ plays a central role, providing efficient feature extraction and encoding of complex 3D scene geometries. This is achieved by recursively applying PointNet++ to nested partitions of the input point set in metric space. Convex decomposition, an important step in the approach, allows transforming complex three-dimensional objects into a set of simpler convex shapes. This simplifies data processing and makes the reconstruction process more manageable. The Transformer then trains the model on these features, allowing for the generation of high-quality reconstructions. It is important to note that the Transformer is used exclusively to determine the position of walls and object boundaries. This combination of technologies allows achieving high accuracy in the reconstruction of 3D models. The main idea of the method is to segment the point cloud into small fragments, which are then restored as polygonal meshes. To restore missing points in point cloud data, a method based on the L1-Median algorithm and local point cloud features is used. This approach can adapt to various geometric structures and correct topological connection errors. The proposed method was compared with several modern approaches and showed its potential in various fields, including architecture, engineering, digitization of cultural heritage, and augmented and mixed reality systems. This underscores its wide applicability and significant potential for further development and application in various fields.

*Keywords:* point cloud, 3D models, neural networks, PointNet++, transformer, convex decomposition