ПРОГРАММИРОВАНИЕ, 2023, № 3, с. 26–36

КОМПЬЮТЕРНАЯ ГРАФИКА И ВИЗУАЛИЗАЦИЯ

УДК 004.92

МЕТОД 3D-РЕКОНСТРУКЦИИ И ОЦИФРОВКИ СЦЕНЫ ДЛЯ СИСТЕМ СМЕШАННОЙ РЕАЛЬНОСТИ

© 2023 г. М. И. Сорокин[a,*] (ORCID: 0000-0001-9093-1690), Д. Д. Жданов[a,] (ORCID: 0000-0001-7346-8155), А. Д. Жданов[a,***] (ORCID: 0000-0002-2569-1982)**

[a]*Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Россия, 197101 Санкт-Петербург, Кронверкский пр., 49*

**e-mail: vergotten@gmail.com*

***e-mail: ddzhdanov@mail.ru*

****e-mail: andrew.gtx@gmail.com*

Поступила в редакцию 09.01.2023 г.
После доработки 16.01.2023 г.
Принята к публикации 20.01.2023 г.

Системы смешанной реальности являются перспективным направлением, открывающим большие возможности для взаимодействия с виртуальными объектами в реальном мире. Как любое перспективное направление смешанная реальность имеет ряд нерешенных проблем. Одна из таких проблем – это формирование естественных условий освещения для виртуальных объектов, а также обеспечение корректного светового взаимодействия виртуальных объектов с реальным миром. Так как виртуальные и реальные объекты находятся в разных пространствах, то обеспечить их корректное взаимодействие является сложной задачей. Для создания цифровых копий объектов реального мира используются инструменты машинного обучения и технологии нейронных сетей. Данные методы успешно применяются в задачах компьютерного зрения для решения проблем ориентации в пространстве и реконструкции окружающей среды. В качестве решения предлагается переместить все объекты в одно информационное пространство – виртуальное. Такое решение позволит снять большую часть проблем, связанных с дискомфортом зрительного восприятия, вызванного неестественным световым взаимодействием объектов реального и виртуального миров. Поэтому основная идея метода заключается в определении объектов физического мира по облакам точек и их замена виртуальными CAD-моделями. То есть семантический анализ сцены и задача классификации объектов с последующим преобразованием в полигональные модели. В данной работе предлагается использование конкурентоспособных нейросетевых архитектур, позволяющих получить современные “state of the art” результаты. Эксперименты проводились на наборах данных “Semantic3D”, “ScanNet” и “S3DIS”, которые на данный момент являются крупнейшими датасетами с наборами облаков точек интерьерных сцен. В качестве метода решения задач семантической сегментации и классификации 3D-облаков точек было решено использовать архитектуру PointNeXt, основанную на PointNet, и применить в процессе обучения современные методы аугментации данных. Для восстановления геометрии был рассмотрен метод дифференциального рендеринга Soft Rasterizer и нейронная сеть “Total3Understanding”.

DOI: 10.31857/S0132347423030056, **EDN:** DEMESA

1. ВВЕДЕНИЕ

За последние несколько десятилетий благодаря новым методам формирования визуальной и звуковой информации взаимодействие между человеком и компьютером претерпело множество изменений. Вслед за мэйнфреймами, ПК и мобильными телефонами следующим революционным компонентом вычислительной техники становится смешанная реальность. Объединение физической и виртуальной реальностей позволяет обеспечить более естественное и интуитивно понятное взаимодействие между людьми, машинами и окружающей их средой.

Системы смешанной реальности становятся популярны как среди потребителей, так и среди компаний. Границы областей применения систем смешанной реальности определяет только наше воображение. Медицина, образование, воссоздание исторических памятников и достопримечательностей, коммуникация и игровая индустрия — это лишь малая часть потенциального использования данных систем и приложений [1–5].

Системы смешанной реальности с технической точки зрения хоть и похожи на системы виртуальной и дополненной реальности, однако имеют существенные отличия. Виртуальный мир

должен восприниматься как единое целое с реальным миром, выглядеть физически корректно и естественно. Отсюда возникает ряд задач, включающих физически-корректное отображение виртуального объекта и его встраивание в реальную среду. В зависимости от типа систем смешанной реальности взаимодействие объектов виртуального и реального миров происходят поразному. В видео-прозрачных системах реальный мир передается на дисплей парой камер высокого разрешения, в то время как в оптически-прозрачных системах реальный мир передается через прозрачную оптику и виртуальный объект является прозрачным, а его видимость достигается за счет многократного увеличения его яркости по сравнению с реальной средой. Так как объекты реального и виртуального миров объединяются в так называемом “спектре смешанной реальности”, отсюда возникает ряд проблем, связанных с неправильным взаимным освещением объектов реального и виртуального миров и разного рода окклюзиями. При невыполнении условий физически-корректного отображения виртуального объекта у пользователя системы смешанной реальности возникает дискомфорт зрительного восприятия. Данный дискомфорт негативно сказывается на состоянии пользователя, что может приводить к головокружениям, потере координации, эффекту “укачивания” и даже травмам [6, 7]. Поэтому устранение дискомфорта зрительного восприятия является первостепенной задачей. Программная реализация сложного физического взаимодействия между реальными и виртуальными объектами является сложной и ресурсоемкой технической задачей, так как объекты находятся в разных пространствах и используют разные средства анализа и отображения. Одним из возможных решений является перевод всех объектов в единое пространство, а поскольку перевод виртуальных объектов в реальный мир невозможен, то предлагается оцифровка реального мира и перевод его в виртуальное пространство.

Основой для построения виртуального аналога реального мира служит :глубина” или RGB-D-изображение. На данный момент по точности получения карт глубин лидируют сканирующие лидары, однако их пространственное разрешение и частота обновления кадров (FPS) довольно низкие по сравнению с рядом других методов, например, по сравнению с ToF (Time-of-flight) камерами, которые используются в смартфонах. Стереокамеры на данный момент являются наиболее дешевым аппаратным способом получения карты глубины сцены. Стереокамеры хорошо работают при солнечном освещении, однако менее эффективны при слабом освещении, что может стать проблемой при работе с интерьерными сценами. Основная сложность получения глубины из стереоизображения заключается в постобработке, а именно в построении карт диспаритета, по которым и вычисляются карты глубин сцены. Данная процедура весьма ресурсоемка и не все мобильные устройства могут позволить себе эту процедуру. Также из недостатков следует отметить плохую работу стереокамер на неконтрастных объектах и необходимость иметь большую базу (расстояние между камерами) для определения глубин удаленных объектов. Еще одно интересное и многообещающее решение это ToF-камеры. Данные камеры измеряют временную задержку отраженного света. Изображения карты глубин получаются хоть и низкого разрешения, однако с высоким FPS. Одним из плюсов ToF-камер является высокая скорость работы, что для многих задач является определяющим условием. ToF-камеры отлично работают как при слабом освещении, так и при полном отсутствии света. Точность получения глубины сцены проигрывает лидарам, однако выигрывает у стереокамер. По сравнению со стереокамерами сложность программного обеспечения у ToF-камер низкая, что делает их отличным кандидатом для использования в приложениях реального времени. Несмотря на то, что большинство рассмотренных методов имеет определенные минусы, современные нейросетевые технологии позволяют нивелировать часть этих недостатков.

2. АНАЛИЗ ТЕКУЩЕГО СОСТОЯНИЯ И ПЕРСПЕКТИВ РАЗВИТИЯ КАМЕР ГЛУБИН

В сфере смешанной реальности в настоящее время растет конкуренция со стороны значимых ИТ-компаний. В качестве примеров можно привести PlayStation VR от Sony, Daydream, Cardboard, Tango и ARCore от Google, Oculus Rift от Facebook, Vive от Valve/HTC HTC и HoloLens от Microsoft, а также Apple (ARKit), Acer, Asus, Dell, HP и Lenovo [8]. Цель MR (Mixed Reality) технологий объединить реальный и виртуальный миры таким образом, чтобы “континуум смешанной реальности” казался единым целым. Техническая сложность заключается в точной регистрации виртуальных и реальных объектов. И MR, и VR (Virtual Reality) требуют отображения компьютерных 3D-изображений в реальном времени, для чего разрабатывают и используют новые технологические решения. В настоящее время MR пытается улучшить методы отслеживания объектов, в первую очередь для мобильных устройств. Технология получает все более широкое распространение в результате постоянного улучшения качества приложений смешанной реальности и соответствующего повышения производительности аппаратного обеспечения. Это создает новые области применения для бизнеса, особенно для промышленных приложений, таких как: удален-

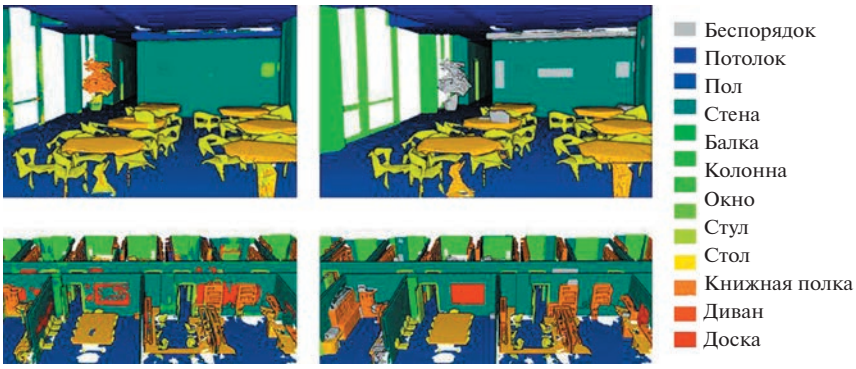

Рис. 1. Пример набора данных S3DIS [11].

Рис. 2. Пример набора данных ScanNe [12].

ная помощь, интерактивное обучение, учебные сценарии и виртуальное прототипирование. Однако существующие решения все еще находятся на стадии экспериментов и часто имеют довольно низкую привлекательность для конечного пользователя и слишком низкую надежность для повседневной эксплуатации, особенно для профессионального использования. Универсальные приложения MR, свободные от дискомфорта зрительного восприятия, отсутствуют на рынке. Кроме того, не хватает и корпоративной поддержки для превращения знаний в инновации. Согласно исследованию Марка Пэллота [9], последние публикации и исследовательские инициативы ясно демонстрируют тенденцию среди предприятий и общественных организаций к открытости в отношении совместного творчества. Для оценки новых идей и концепций идеально подходят иммерсивные и совместные среды (ICE), основанные на дополненной или смешанной реальности. Однако, необходимо разработать методологии и инструменты более сложные, чем те, которые доступны сейчас.

Чтобы быстро адаптироваться к расположению и ориентации дисплея в среде MR, данные должны генерироваться в режиме реального времени с частотой обновления примерно 60 кадров в секунду. Сложность демонстрируемого виртуального контента оказывает значительное влияние на частоту обновления. В связи с тем, что многие системы смешанной реальности являются мобильными и имеют малую вычислительную мощность, данные САПР, результаты сканирования и результаты моделирования часто оказываются слишком сложны для отображения в реальном времени.

Эффективное взаимодействие с данными, полученными от окружающей среды, представляет

Таблица 1. Основные метрики оценки результатов при работе с 3D и облаками точек

Метрика	Формула
Accuracy	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$
mACC	$mACC = \frac{1}{C}\sum_{c=1}^{C} Accuracy_c$
Precision	$Precision = \frac{\mathrm{TP}}{2\mathrm{TP} + \mathrm{FN}}$
Recall	$Recall = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}$
F1-Score	$F_1 = 2 \times \frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}$
IoU	$IoU_i = \frac{I_{i,i}}{\sum_{c=1}^{C}(I_{i,c} + I_{c,i}) - I_{i,i}}$
mIoU	$mIoU = \frac{1}{C}\sum_{c=1}^{C} IoU_i$
EPE	$EPE = \|s\hat{}f - sf\|_2$

собой дополнительную проблему. В отличие от настольных компьютерных систем, для которых клавиатура и мышь являются стандартными устройствами ввода, и мобильных устройств, взаимодействующих при помощи тактильных прикосновений, устройства ввода в технологиях MR отнюдь не стандартизированы. Распознавание речи и управление жестами на данный момент являются основными методами ввода в системах смешанной реальности. Жесты часто используются, например, для выбора пункта в меню. Однако, управление жестами не лучший вариант, если требуется сложный ввод текста или точное размещение виртуальных объектов в пространстве. Отсутствие обратной связи в виде тактильных ощущений и усталость пользователя – также два фактора, которые необходимо учитывать при оценке дискомфорта [10]. Все это представляет серьезный риск для пригодности и развития технологии.

В данной работе предлагается метод, основанный на использовании нейронных сетей для классификации и сегментации объектов сцены, представленных в виде облака точек и методов дифференцируемого рендеринга для восстановления геометрии объектов, чьи аналоги не были найдены в базах данных CAD-объектов.

Таблица 2. Сравнение моделей в задаче семантической сегментации на наборе данных S3DIS

Модель	**mIoU**	**m(Acc)**	**o(Acc)**
WindowNorm + + StratifiedTransformer	77.6	85.8	91.7
PointMetaBase-XXL	77.0	–	91.3
PointNeXt-XL	74.9	83.0	90.3
DeepViewAgg	74.7	83.8	90.1
PointTrans-former+GAM	74.4	83.2	90.6
RepSurf-U	74.3	82.6	90.8
PointNeXt-L	73.9	82.2	89.9
PointTransformer	73.5	81.9	90.2
CBL	73.1	79.4	89.6

3. НАБОРЫ ДАННЫХ ИНТЕРЬЕРНЫХ СЦЕН И МЕТРИКИ ОЦЕНКИ

Существует множество наборов данных с облаками точек, используемых для различных задач, конкретно в нашем случае необходимо сделать акцент именно на наборы данных интерьерных сцен. Из таких датасетов следует отметить S3DIS и ScanNet.

Набор данных Stanford Large-Scale 3D Indoor Spaces (S3DIS) [11] состоит из 5 крупномасштабных внутренних сцен из трех зданий. Они отличаются по архитектуре, внешнему виду и стилю. Данный набор промаркирован 13 классами, что включает в себя конструктивные элементы (пол, стена и т. д.) и обычную мебель. Пример набора данных S3DIS представлен на рис. 1.

ScanNet [12] это большой набор видеоданных (2.5 миллионов кадров), полученных из более чем 1000 сканирований, аннотированных реконструкцией поверхности, семантической сегментацией и положениями камеры. Пример набора данных ScanNet представлен на рис. 2.

Для оценки качества работы алгоритмов и обученных моделей используются специальные метрики оценки, представленные в табл. 1. Данные метрики широко используются как в различных ML-задачах, так и при работе с облаками точек. От показателей данных метрик зависит то, насколько хорошо алгоритм справляется с поставленной задачей, и то, какие выводы можно получить. Для задач сегментации обычно используются метрики accuracy или mIoU, а в задачах детектирования – mIoU, accuracy, precision и recall. Для сопоставления 3D-моделей в сцене могут быть использованы кривые ROC (производные от precision и recall) [13].

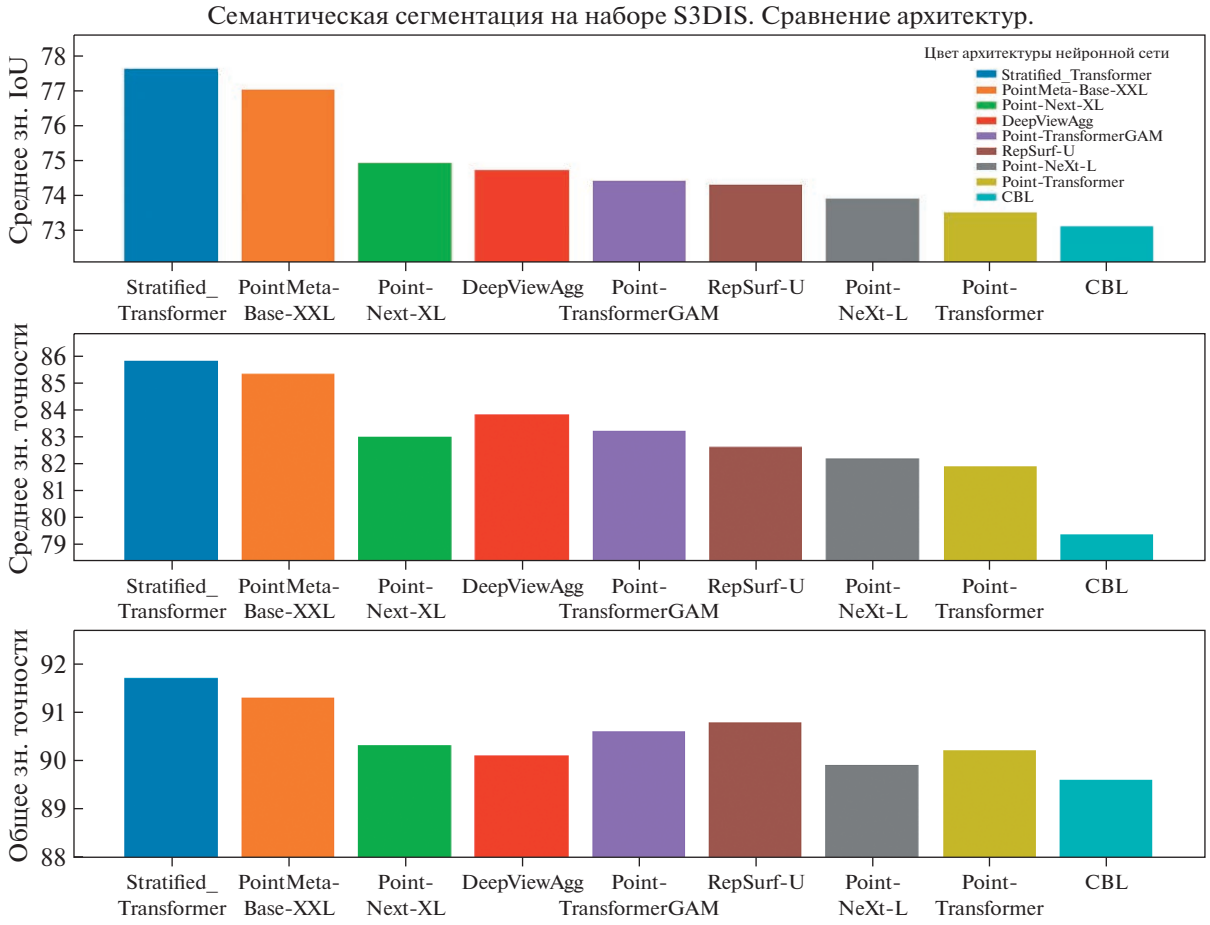

Рис. 3. Сравнение точности классификации нейронных сетей в задачах сегментации.

Accuracy (точность) – это показатель, который обычно описывает, как модель работает на всех классах. mACC (средняя точность) может быть использована в тех случаях, когда классы не сбалансированы. Precision можно рассматривать как меру качества, а recall – как меру количества. Показатель F1 можно интерпретировать как среднее гармоническое между precision и recall, где показатель F1 достигает своего наилучшего значения при 1, а наихудшего значения при 0. IoU пересечение определяется между предсказанной маской и ground truth. mIoU – среднее IoU на всех классах. End point error (EPE) используется в задачах вычисления оптического потока.

Таблица 3. Сравнение моделей в задаче классификации 3D-облака точек на наборе ScanObjectNN

Модель	m(Acc)	o(Acc)
I2P-MAE	90.11	–
ULIP + PointNeXt	89.7	88.6
ULIP + PointMLP	89.4	88.5
P2P	89.3	–
PointNeXt+Local	88.6	87.4
PointNeXt+GAM	88.4	–
PointNeXt+HyCoRe	88.3	87.0
ACT	88.21	–
PointNeXt	88.2	86.8

4. АРХИТЕКТУРА POINTNEXT ДЛЯ РАБОТЫ С 3D-ОБЛАКАМИ ТОЧЕК

На текущий момент существует большое количество архитектур нейронных сетей для работы с трехмерными данными. Нейронные сети могут работать практически с любым типом 3D-информации: как с облаками точек, так и с вокселями и полигональными сетками.

В данной работе рассмотрена архитектура PointNeXt [14], что является следующей версией таких архитектур как PointNet и PointNet++. Как утверждают авторы, данная архитектура еще не исчерпала себя и, благодаря современным мето-

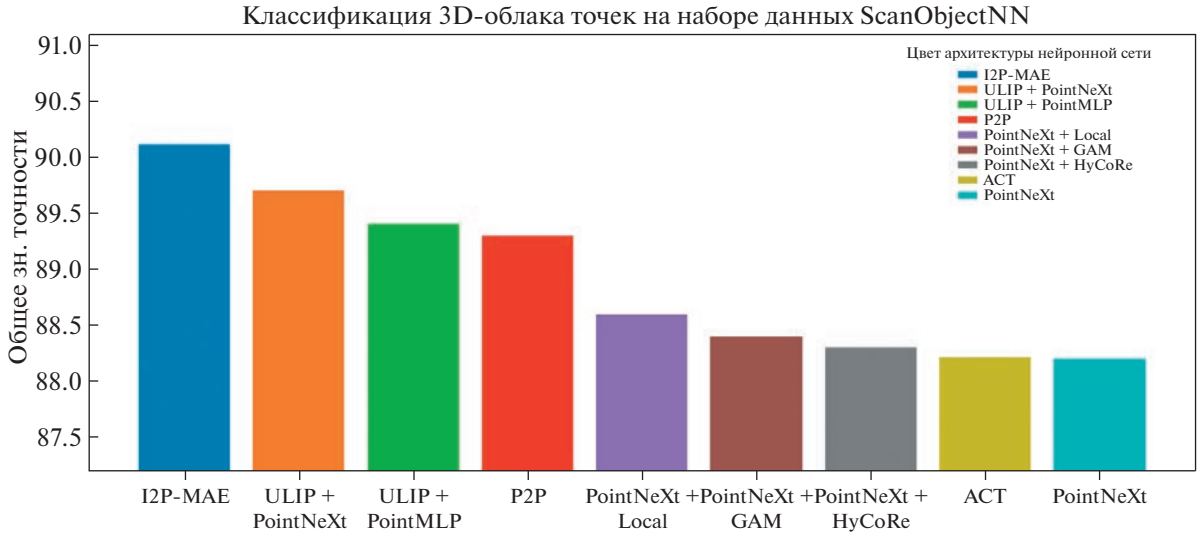

Рис. 4. Сравнение точности классификации нейронных сетей в задачах классификации 3D-облака точек.

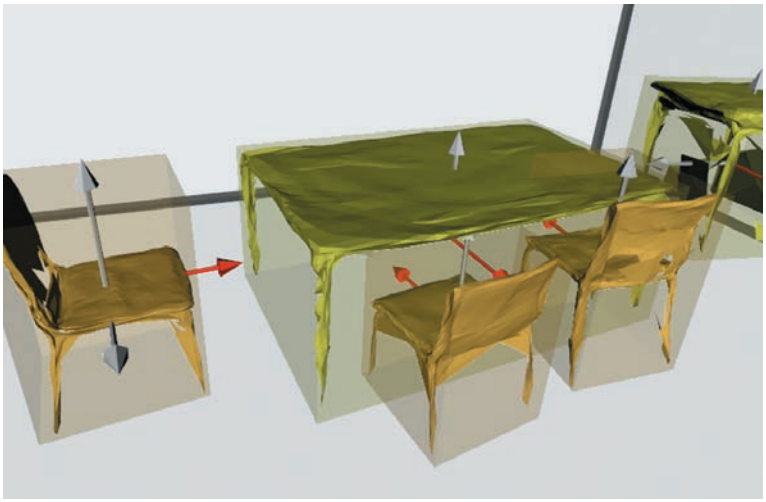

Рис. 5. Выделение объектов сцены в пространстве с использованием Total3DUnderstanding.

дам обучения и аугментации данных, позволяет достичь современных SOTA (State Of The Art) результатов. В задачах семантической сегментации PointNeXt хоть и не сильно, но уступает таким архитектурам, как StratifiedTransformer и PointMeta-Base-XXL (рис. 3, табл. 2). Однако, PointNeXt также показывает отличные результаты в задачах 3D-классификации по облаку точек (рис. 4, табл. 3).

Для эффективного масштабирования модели авторы добавили в PointNet++ разделяемые многослойные перцептроны (MLPs) [15], инвертированный “боттлнек” (inverted bottleneck design) [16] и остаточные связи (residual connections) [17]. Только при пересмотре методов обучения общая точность OA (Overall Accuracy) на наборе ScanObjectNN увеличивается на 8.2% (с 77.9 до 86.1%), создавая новую SOTA без внесения каких-либо изменений в архитектуру. Показатель mIoU (mean Intersection Over Union), оцененный на всех регионах с помощью 6-кратной перекрестной валидации на наборе S3DIS, увеличивается на 13.6% (с 54.5 до

Таблица 4. Общая точность, средняя точность и mIoU на тестовой сцене

Общ. точность (Overall Accuracy)	Средн. точность (Mean Accuracy)	Средн. IoU (Mean IoU)
90.94	77.13	76.55

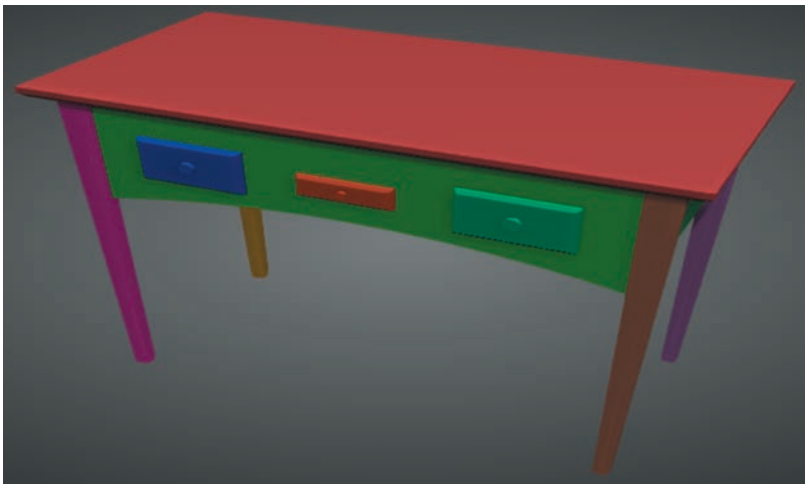

Рис. 6. Пример сегментации частей объекта.

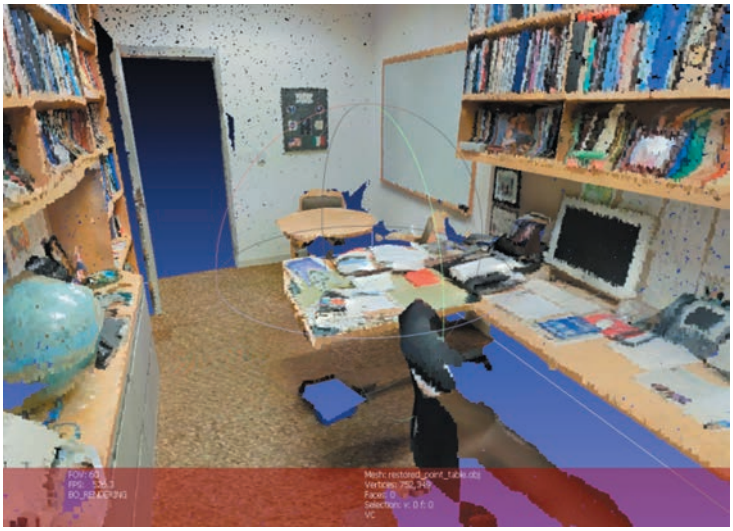

Рис. 7. Скан интерьерного помещения.

68.1%), опережая несколько современных разработок, появившихся после PointNet++, таких как PointCNN [18] и DeepGCN [19]. Производительность нейронной сети также зависит от методов оптимизации, таких как функции потерь, оптимизаторов, планировщиков скорости обучения и гиперпараметров. Благодаря развитию теории машинного обучения современные нейронные сети можно обучать с помощью более совершенных оптимизаторов (например, AdamW [20] по сравнению с Adam [21]) и более продвинутых функций потерь (CrossEntropy со сглаживанием меток [22]).

Производительность в задачах классификации и сегментации повышается благодаря увеличению объема данных (аугментации). В ScanObjectNN повторная выборка точек улучшает производительность на 2.5% OA. Результат сегментации улучшается на 1.1% mIoU, когда в качестве входных данных используется полная сцена в отличие от данных, дискретизированных по блокам или сферам.

5. МЕТОД 3D-РЕКОНСТРУКЦИИ И ОЦИФРОВКИ СЦЕНЫ

Для решения поставленной задачи восстановления геометрии реального окружения, наблюдаемого стереокамерами системы MR, представлен алгоритм, который основывается на нейронных сетях “PointNeXt” и “Total3DUnderstanding” [23].

1. PointNeXt используется для классификации облака точек, в то время как Total3DUnderstanding – для восстановления границ и положения объектов в пространстве сцены (рис. 5). Помимо

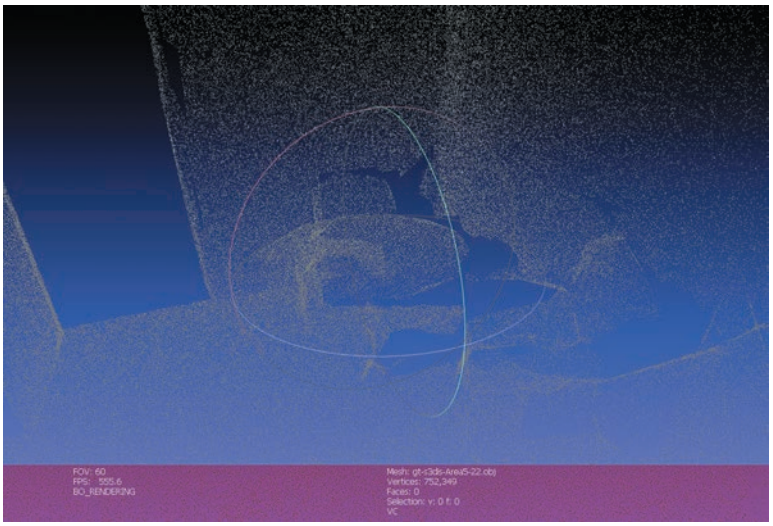

Рис. 8. Облако точек интерьерного помещения.

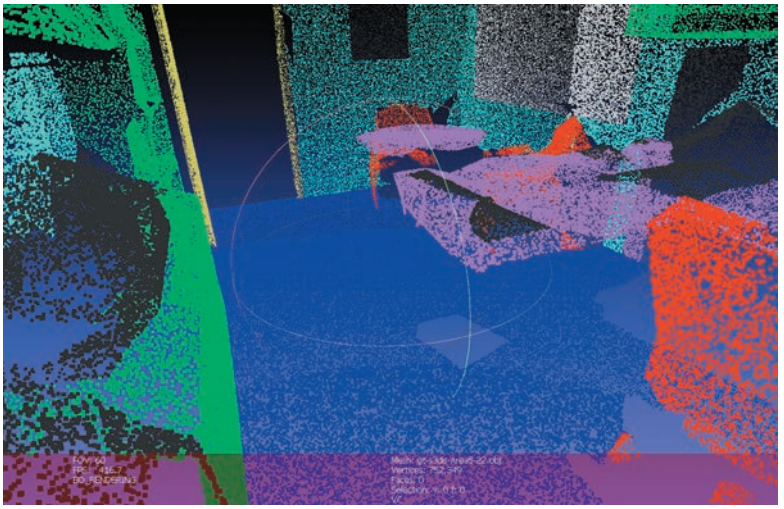

Рис. 9. Сегментированное PointNeXt облако точек.

LEN компонента (Layout Estimation Network) Total3DUnderstanding включает в себя такие модули, как ODN (Object Detection Network) и MGN (Mesh Generation Network).

2. После классификации облаков точек по классам объектов определяются их составные части (столешница, ножки, ящики стола), т.е. выделяются их сегменты (рис. 6).

3. Необходимо составить достаточно большую базу данных с CAD-моделями (либо использовать готовую, например ShapeNetCore [24]). Данная база данных должна иметь подробную аннотацию и описание составных частей объекта для максимально быстрого поиска аналога и замены им реального объекта сцены.

4. Если по облаку точек не удалось составить полную картину объекта и его частей, используются средства дифференциального рендеринга, где в качестве целевого объекта выбирается объект, максимально похожий по описанию классифицированных частей на реальный объект сцены.

5. Из базы данных извлекается модель и помещается на место реального объекта в виртуальном пространстве сцены.

6. РЕЗУЛЬТАТЫ

В качестве тестовой сцены для эксперимента был взят скан интерьерного помещения из набора данных 3SDIS, который не был использован при обучении (рис. 7).

Таблица 5. IoU по всем предсказанным классам

N класса	Название класса	IoU
1	Беспорядок (Clutter)	93.68
2	Потолок (Ceiling)	98.58
3	Пол (Floor)	85.26
4	Стена (Wall)	71.08
5	Балка (Beam)	42.27
6	Колонна (Column)	60.59
7	Окно (Window)	70.93
8	Дверь (Door)	84.44
9	Кресло (Chair)	92.41
10	Стол (Table)	80.35
11	Книжная полка (Bookcase)	78.08
12	Диван (Sofa)	76.59
13	Доска (Board)	60.98

Соответствующее ему облако точек представлено на рис. 8.

Результат работы нейронной сети PointNeXt представлен на рис. 9.

Кластеры точек, предсказанные к разным классам, для наглядности выделены разным цветом. В табл. 4 представлены значения общей и средней точности тестовой сцены, а в табл. 5 значения IoU для каждого класса объектов и их наименование.

После завершения классификации выполняется построение полигональных сеток по облакам точек. Для этого используется алгоритм реконструкции поверхности Пуассона. Для случаев, когда восстановленной полигональной модели оказывается достаточно для определения частей объекта, осуществляется замена облака на найденную в базе данных CAD-модель (рис. 10). Данная модель масштабируется и размещается в области пространства, занимаемой облаком точек. Для случаев, когда сначала необходимо восстановить отсутствующие части объекта, применяется дифференцируемый рендеринг SoftRasterizer.

Целевая модель – это конечная модель, которую мы хотим получить при дифференцируемом рендеринге. Другими словами, это эталонная модель, к которой мы стремимся деформировать изначальную, используя функции потерь. Так как в исходном облаке точек есть отсутствующие элементы (рис. 12, исходное облако точек), требуется по текстовому описанию найти наиболее подходящую модель или скомпоновать ее из частей объектов. В данном примере в качестве целевой модели была выбрана модель из набора ShapeNetCore, которая максимально похожа по описанию присутствующих частей (круглая спинка стула, ручки).

Полученная после дифференцируемого рендеринга модель слишком грубая и содержит большое количество артефактов (рис. 12, полигональная сетка после дифференцируемого рендеринга). Для финальной сцены такую модель использовать нельзя, однако можно найти наиболее подходящий аналог из базы данных. Поэтому деформированная модель заменяется CAD-моделью из базы данных (рис. 12, CAD-модель). В процессе выполнения дифференцируемого рендеринга используются следующие метрики и функции потерь: chamfer loss, edge loss, normal loss, laplacian loss (рис. 11), где chamfer loss – расстояние между прогнозируемой (деформированной) и целевой моделью, edge loss минимизирует длину ребер в прогнозируемой модели, normal loss – согласованность нормалей соседних граней, а laplacian loss является регуляризатором.

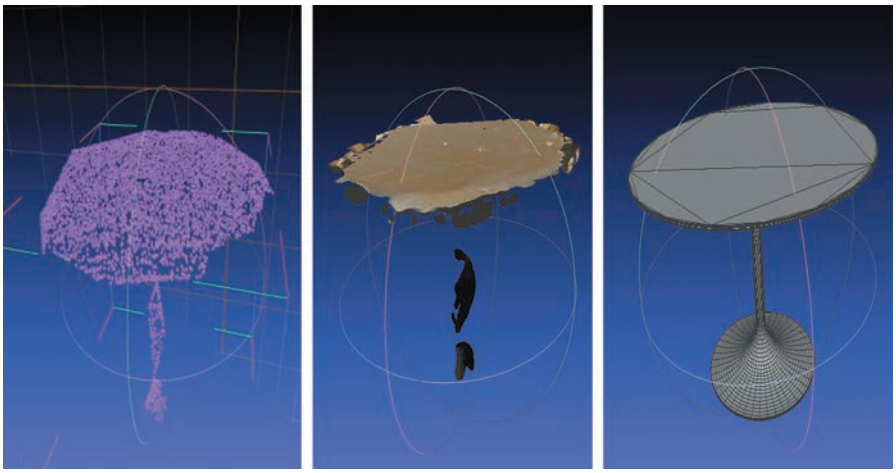

Рис. 10. Облако точек, полигональная сетка и CAD аналог (слева направо).

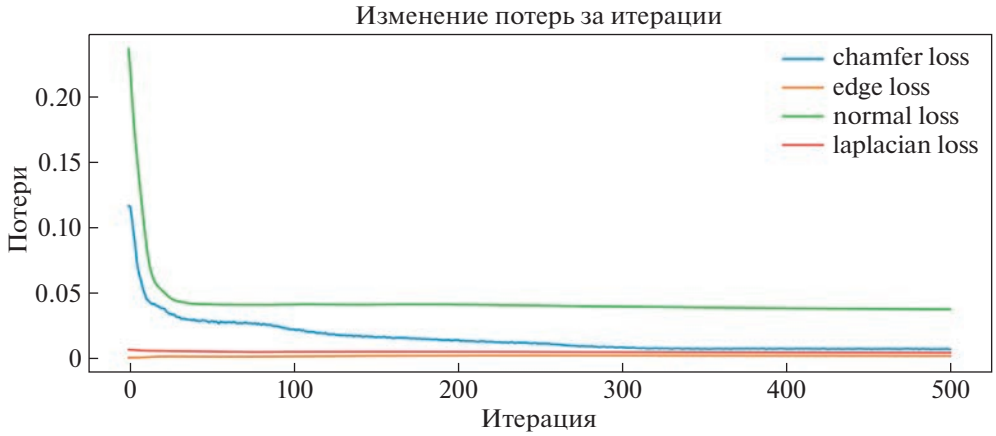

Рис. 11. Изменение функций потерь при дифференцируемом рендеринге на тестовой модели.

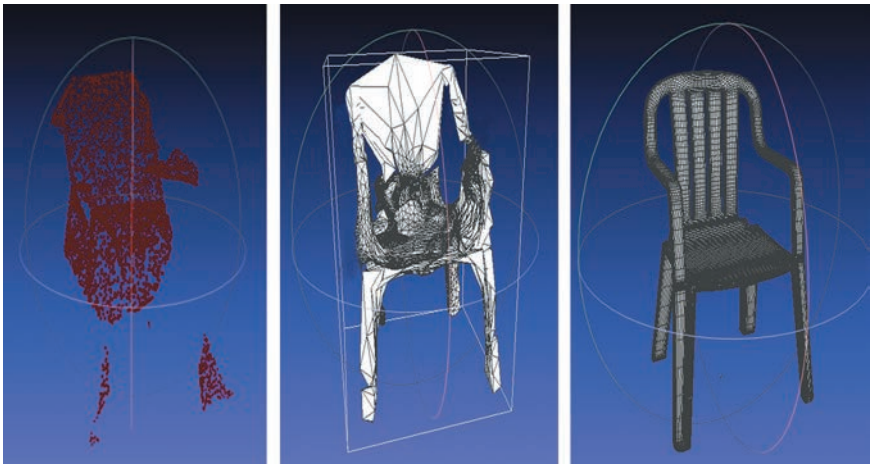

Рис. 12. Исходное облако точек, полигональная сетка после дифференцируемого рендеринга и CAD модель (слева-направо).

7. ЗАКЛЮЧЕНИЕ

Системы смешанной реальности представляют собой перспективное направление, обеспечивающее интерактивное взаимодействие человека с объектами реального и виртуального миров одновременно. Однако, проблема корректного взаимодействия виртуальных объектов с физическим миром полностью не решена, и это замедляет процесс внедрения данных систем в повседневную жизнь. В данной работе был представлен метод реконструкции реальной сцены для систем смешанной реальности с использованием облаков точек и архитектур нейронных сетей PointNeXt и Total3DUnderstanding. Метод основан на замене реального объекта его виртуальным аналогом, что позволяет устранить ряд источников дискомфорта зрительного восприятия, связанных со световым взаимодействием объектов реального и виртуального миров. Нейронная сеть PointNeXt показывает достаточно хороший результат в задачах классификации облака точек, достигая SOTA результатов, в то время как Total3DUnderstanding решает важные задачи по определению границ и ориентации объектов в пространстве сцены. В данной работе использовался набор данных 3SDIS, который содержит 13 классов. В дальнейшем планируется использование набора данных ScanNet200, который уже содержит 200 классов различных объектов интерьерного помещения. Также планируется создать большую базу данных интерьерных объектов с подробным описанием, чтобы обеспечить эффективный поиск CAD-модели. Для поиска модели по заданным текстовым параметрам и аннотациям перспективно смотрится реализация нейронной сети по типу трансформера.

ФИНАНСИРОВАНИЕ

Работа была выполнена при финансовой поддержке Российского научного фонда, грант № 22-11-00145.

СПИСОК ЛИТЕРАТУРЫ

1. *Dhaval S.* Critical review of mixed reality integration with medical devices for patientcare // International Journal for Innovative Research in Multidisciplinary Field. 2022. V. 8. Issue 1. https://doi.org/10.2015/IJIRMF/202201017
2. *Maas M.J., Hughes J.M.* Virtual, augmented and mixed reality in K-12 education: a review of the literature // Technology, Pedagogy and Education. 2020. V. 29. Issue 2. https://doi.org/10.1080/1475939X.2020.1737210
3. *Evangelidis K., Sylaiou S., Papadopoulos T.* Mergin'mode: Mixed reality and geoinformatics for monument demonstration // Applied Sciences. 2020. V. 10. № 11. P. 3826.
4. *Piumsomboon T., Lee G.A., Hart J.D., Ens B., Lindeman R.W., Thomas B.H., Billinghurst M.* Mini-me: An adaptive avatar for mixed reality remote collaboration / In Proceedings of the 2018 CHI conference on human factors in computing systems. 2018. P. 1–13.
5. *Miedema N.A., Vermeer J., Lukosch S., Bidarra R.* Superhuman sports in mixed reality: The multi-player game League of Lasers / In 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). IEEE, 2019. P. 1819–1825.
6. *Guna J., Gersak G., Humar I.* Virtual Reality Sickness and Challenges Behind Different Technology and Content Settings // Mobile Networks and Applications. 2020. V. 25. P. 1436–1445. https://doi.org/10.1007/s11036-019-01373-w
7. *Saredakis D., Szpak A., Birckhead B., Keage H.A., Rizzo A., Loetscher T.* Factors associated with virtual reality sickness in head-mounted displays: a systematic review and meta-analysis // Frontiers in human neuroscience. 2020. V. 14. P. 96.
8. *Moser T., Hohlagschwandtner M., Kormann-Hainzl G., Pölzlbauer S., Wolfartsberger J.* Mixed reality applications in industry: challenges and research areas / In International Conference on Software Quality. Cham.: Springer, 2019. P. 95–105.
9. *Pallot M., Fleury S., Poussard B., Richir S.* What are the Challenges and Enabling Technologies to Implement the Do-It-Together Approach Enhanced by Social Media, its Benefits and Drawbacks? // Journal of Innovation Economics Management. 2022. I132-XLII.
10. *Guo J., Weng D., Zhang Z., Liu Y., Duh H.B., Wang Y.* Subjective and objective evaluation of visual fatigue caused by continuous and discontinuous use of HMDs // Journal of the Society for Information Display. 2019. V. 27, № 2. P. 108–119.
11. *Armeni I., Sener O., Zamir A.R., Jiang H., Brilakis I., Fischer M., Savarese S.* 3D semantic parsing of large-scale indoor spaces / In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016. P. 1534–1543.
12. *Dai A., Chang A.X., Savva M., Halber M., Funkhouser T., Nießner M.* ScanNet: Richly-annotated 3D reconstructions of indoor scenes / In Proc. Computer Vision and Pattern Recognition (CVPR), IEEE, 2017.
13. *Haoming L., Humphrey S.* Deep Learning for 3D Point Cloud Understanding: A Survey // Computer Vision and Pattern Recognition. 2020. https://doi.org/10.48550/arXiv.2009.08920
14. *Qian G., Li Y., Peng H., Mai J., Hammoud H.A., Elhoseiny M., Ghanem B.* PointNeXt: Revisiting PointNet++ with Improved Training and Scaling Strategies. arXiv preprint arXiv:2206.04670. 2022.
15. *Qian G., Hammoud H., Li G., Thabet A., Ghanem B.* Assanet: An anisotropicseparable set abstraction for efficient point cloud representation learning // Advances in Neural Information Processing Systems (NeurIPS). 2021. P. 34.
16. *Sandler M., Howard A., Zhu M., Zhmoginov A., Chen L.C.* Mobilenetv2: Inverted residuals and linear bottlenecks / In Proceedings of the IEEE/CVF Conference on Computer Visionand Pattern Recognition (CVPR). 2018. P. 4510–4520.
17. *He K., Zhang X., Ren S., Sun J.* Deep residual learning for image recognition / In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.
18. *Li Y., Bu R., Sun M., Wu W., Di X., Chen B.* Pointcnn: Convolution on X-transformed points // Advances in Neural Information Processing Systems (NeurIPS), 2018.
19. *Li G., Muller M., Thabet A., Ghanem B.* Deepgcns: Can gcns go as deep as cnns? / In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2019. P. 9267–9276.
20. *Loshchilov I., Hutter F.* Decoupled weight decay regularization / In International Conference on Learning Representations (ICLR). 2019.
21. *Diederik P. Kingma, Jimmy Ba.* Adam: A method for stochastic optimization / In International Conference on Learning Representations (ICLR). 2015.
22. *Szegedy C., Vanhoucke V., Ioffe S., Shlens J., Wojna Z.* Rethinking theinception architecture for computer vision / In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2016.
23. *Nie Y., Han X., Guo S., Zheng Y., Chang J., Zhang J.J.* Total3dunderstanding: Joint layout, object pose and mesh reconstruction for indoor scenes from a single image / In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. P. 55–64.
24. *Kulikajevas A., Maskeliūnas R., Damaševičius R., Misra S.* Reconstruction of 3D object shape using hybrid modular neural network architecture trained on 3D models from ShapeNetCore dataset // Sensors. 2019. V. 19. № 7. P. 1553.