МИНОБРНАУКИ РОССИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра дискретной математики и
информационных технологий

ПРЕОБРАЗОВАНИЕ 2D В 3D С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВЫХ АРХИТЕКТУР ПОСТРОЕНИЯ КАРТЫ ГЛУБИНЫ

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 421 группы
направления 09.03.01 — Информатика и вычислительная техника
факультета КНиИТ
Васильева Егора Андреевича

Научный руководитель
д.экон.н., профессор _____ Л. В. Кальянов

Заведующий кафедрой
к. ф.-м. н., доцент _____ Л. Б. Тяпаев

Саратов 2026

ВВЕДЕНИЕ

Актуальность темы. В современном мире цифровых технологий и визуальных искусств преобразование изображений из 2D в 3D становится все более востребованной задачей. Это связано с широким спектром применения трехмерных моделей в киноиндустрии, видеоиграх, виртуальной и дополненной реальности, архитектуре, медицине и многих других областях. Традиционные методы получения трехмерных моделей — ручное моделирование в 3D-редакторах, фотограмметрия, использование стереопар — либо требуют высокой квалификации и значительных временных затрат, либо нуждаются в дорогостоящем оборудовании (LiDAR) или нескольких снимках.

Современные методы глубинного обучения предлагают принципиально иной подход: нейросетевые архитектуры для монокулярной оценки глубины позволяют «додумывать» третье измерение по одному RGB-изображению, извлекая статистические закономерности из больших массивов данных. Однако существующие готовые сервисы (Meshy, Polycam) имеют ограничения по функционалу и не позволяют гибко настраивать процесс обработки. В связи с этим актуальной является задача разработки программного комплекса, объединяющего методы оценки глубины и 3D-реконструкции в сквозную вычислительную цепочку.

Цель выпускной квалификационной работы — разработка программного комплекса с использованием библиотек глубокого обучения и компьютерного зрения для преобразования двумерных изображений в трехмерные модели.

Поставленная цель определила следующие задачи:

1. Провести анализ классических методов создания 3D-моделей на основе 2D-изображений;
2. Рассмотреть устройство и принципы работы LiDAR и RGB-D камер как источников данных о глубине сцены;
3. Изучить архитектуры нейронных сетей, применяемые для монокулярной оценки глубины и обработки облаков точек;
4. Рассмотреть существующие программные средства и готовые сервисы для проведения преобразований изображений из 2D формата в 3D;
5. Реализовать программный комплекс для преобразования изображения из 2D формата в 3D модель.

Методологические основы преобразования 2D в 3D представлены в работах Ю. В. Войтешонок, А. П. Кирпичникова, А. Антонова, А. Ю. Черноока, А. Р. Попова.

Теоретическая значимость. В работе систематизированы современные методы монокулярной оценки глубины и проведен обоснованный выбор нейросетевой модели для дообучения в условиях ограниченных вычислительных ресурсов. Предложенный и реализованный сквозной пайплайн преобразования 2D-изображений в 3D-модели может быть использован для дальнейших исследований в области автоматизированной 3D-реконструкции по одному изображению.

Практическая значимость. Разработанный программный комплекс позволяет автоматизировать процесс преобразования двумерных изображений в трехмерные модели, может быть использован в образовательных целях, при разработке прототипов систем виртуальной и дополненной реальности, а также в области трехмерного моделирования интерьеров.

Структура и объем работы. Выпускная квалификационная работа состоит из введения, 4 разделов, заключения, списка использованных источников и приложения. Общий объём работы — 59 страниц, включая 18 рисунков, список использованных источников информации — 22 наименования.

КРАТКОЕ СОДЕРЖАНИЕ РАБОТЫ

Первый раздел «Теоретические основы преобразования 2D-изображений в 3D-модели» посвящен обзору классических методов получения трехмерной информации из изображений, а также рассмотрению аппаратных средств прямого измерения глубины сцены.

В подразделе 1.1 приведена историческая справка, описывающая эволюцию представлений о трёхмерных объектах от геометрических построений Евклида и прямоугольной системы координат Декарта до первых компьютерных методов 3D-моделирования и визуализации.

В подразделе 1.2 приведено описание шести классических методов преобразования 2D в 3D: ручное моделирование в 3D-редакторах (Blender), создание карты глубины по стереопаре, реконструкция из нескольких изображений, фотограмметрия с использованием алгоритма Левенберга-Марквардта, стереоскопическое 3D-преобразование для киноиндустрии и визуальная имитация объема с помощью бликов и теней.

В подразделе 1.3 рассмотрены аппаратные методы измерения глубины — LiDAR и RGB-D камеры. Технология LiDAR (Light Detection and Ranging) основана на измерении времени пролета светового импульса. Расстояние d до точки определяется по формуле:

$$d = \frac{c \cdot \Delta t}{2}$$

где c — скорость света, Δt — время между излучением и регистрацией отражённого сигнала. Деление на 2 учитывает двукратное прохождение расстояния (туда и обратно).

Современные LiDAR-системы включают лазерный излучатель, фотоприемник (APD или SPAD-матрицы), высокоточный таймер (TDC — Time-to-Digital Converter), систему сканирования (механическую или твердотельную) и блок синхронизации. Измерение расстояния в LiDAR может осуществляться несколькими методами: импульсный Time-of-Flight, фазовый метод и FMCW LiDAR (Frequency-Modulated Continuous Wave).

Альтернативным и более доступным решением являются RGB-D камеры, которые совмещают получение цветовой информации и глубины в единой системе координат. Формально результат работы RGB-D камеры — это пара $(I(x,y), D(x,y))$, где $I(x,y)$ — цвет, а $D(x,y)$ — глубина в пикселе. Координаты точки восстанавливаются через параметры камеры:

$$X = \frac{(x - c_x) \cdot D(x,y)}{f_x}, \quad Y = \frac{(y - c_y) \cdot D(x,y)}{f_y}, \quad Z = D(x,y)$$

где f_x, f_y — фокусные расстояния, c_x, c_y — координаты главной точки. Существует три технологии получения глубины: стереоскопическое зрение ($Z = \frac{f \cdot B}{d}$, где B — базис между камерами, d — смещение), структурированная подсветка и Time-of-Flight камеры.

RGB-D камеры часто выступают в качетсве инструмента для сбора эталонных карт глубины при создании датасетов для обучения нейросетей.

Второй раздел «Методы глубокого машинного обучения для оценки глубины сцены» посвящен эволюции нейросетевых архитектур и описанию конкретных моделей, используемых в работе.

В подразделе 2.1 рассмотрена эволюция от свёрточных нейронных се-

тей (CNNs) к трансформерам. Отмечено фундаментальное ограничение CNN — постепенное расширение рецептивного поля, что затрудняет установление глобальных связей между удалёнными областями изображения. Vision Transformer (ViT) решает эту проблему за счёт механизма самовнимания, позволяющего учитывать вклад всех элементов изображения уже на ранних этапах обработки.

В подразделе 2.2 приведено описание гибридных подходов, в частности архитектуре Dense Prediction Transformer (DPT), которая легла в основу третьей версии MiDaS. DPT сохраняет пространственную информацию на разных уровнях трансформера и объединяет многоуровневые признаки для восстановления карты глубины. Существуют две основные версии: DPT-Large на чистом ViT и DPT-Hybrid с добавлением свёрточной сети перед трансформером.

В подразделе 2.3 подробно рассмотрено семейство моделей MiDaS (Mixed Dataset training for Depth and Segmentation). Важной особенностью данной нейронной сети является ее предобучение на 10–12 разнородных датасетах (ReDWeb, DIML, Movies, MegaDepth, WSVD, TartanAir и других) с использованием масштабно-инвариантной функции потерь (scale-and-shift-invariant loss). Эта функция позволяет сети игнорировать различия в абсолютных масштабах между датасетами и фокусироваться на относительной структуре глубины.

В подразделе 2.4 представлена специализированная архитектура GLPN (Global-Local Path Network) для интерьерных сцен. В качестве основы GLPN использует SegFormer — иерархическую архитектуру mix-Transformer, специально разработанную для задач, где каждому пикселю изображения нужно поставить в соответствие некоторое значение (например, глубину). Отличительной особенностью GLPN является легковесный декодер, построенный вокруг модуля Selective Feature Fusion Module (SFFM), который избирательно объединяет локальные признаки разного масштаба с глобальным потоком информации. Также авторы предложили приём аугментации данных под названием Vertical CutDepth: изображение случайным образом разрезается по вертикали на несколько сегментов, и эти сегменты переставляются местами, а карта глубины переставляется аналогичным образом. Это вынуждает сеть учиться определять глубину на основе семантических и текстурных призна-

ков, а не полагаться на простую закономерность «чем выше объект на изображении, тем он дальше».

Третий раздел «Нейросетевые методы постобработки и построения 3D-структур» включает описание методов обработки облаков точек и программной экосистемы Python для 3D-реконструкции.

В подразделе 3.1 рассмотрена архитектура PointNet, решающая проблему неупорядоченности облака точек. Формально облако точек представляется как множество $P = \{p_i \in \mathbb{R}^3 \mid i = 1, \ldots, N\}$, где каждая точка $p_i = (x_i, y_i, z_i)$, а также может быть дополнена атрибутами (цвет, нормаль, интенсивность). Для каждой точки применяется отображение $h : \mathbb{R}^3 \to \mathbb{R}^K$, после чего глобальный дескриптор вычисляется как:

$$g(P) = \max_{i=1,\ldots,N} h(p_i)$$

Операция max pooling является симметричной функцией, что гарантирует инвариантность к перестановке точек. Однако глобальное агрегирование приводит к потере локальной структуры, поэтому была предложена архитектура PointNet++, которая вводит иерархическую обработку, включающую выбор опорных точек, построение локальных окрестностей и извлечение локальных признаков. Также описаны трансформерные методы и диффузионные модели для 3D-данных.

В подразделе 3.2 описаны программные средства, используемые на разных этапах преобразования 2D в 3D на языке Python. Описывается библиотека PyTorch, которая обеспечивает выполнение тензорных операций на GPU, автоматическое дифференцирование и построение нейросетевых моделей. Отмечается библиотека Hugging Face Transformers, предоставляющая унифицированный интерфейс для загрузки предобученных моделей (в частности, GLPN) и их предобработки. Также рассматривается библиотека Open3D, предназначенная для работы с трёхмерными данными: в ней описываются структуры данных для хранения облаков точек и изображений глубины, алгоритмы фильтрации выбросов, оценка нормалей поверхности и построение полигональной сетки методом Пуассона.

Четвёртый раздел «Практическая часть» содержит результаты экспериментального исследования готовых инструментов и реализацию собственного программного комплекса для преобразования 2D в 3D.

В подразделе 4.1 протестированы готовые сервисы: dzine.ai (создание 3D-эффекта за счёт света и теней), Meshy (генерация 3D-моделей из текста или изображения) и Polycam (фотограмметрия с помощью смартфона). Продемонстрирован полный цикл фотограмметрической реконструкции фарфоровых статуэток слоников с последующей доработкой в Blender.

В подразделе 4.2 описано дообучение модели MiDaS_small на датасете NYU Depth V2. Проведена предобработка данных: транспонирование осей, ресайз до 256×256 пикселей, нормализация RGB в диапазон $[0, 1]$ и карт глубины — делением на максимальное значение. Обучение выполнено с использованием оптимизатора Adam (learning rate 10^{-4}) и масштабно-инвариантной функции потерь на 50 эпохах. Результат предсказания карты глубины представлен на рисунке 1.

Рисунок 1 – Предсказанная карты глубины с помощью MiDaS

Далее описан инференс модели GLPNForDepthEstimation (vinvino02/glpn-nyu), которая заранее была обучена на используемом датасете. Предобработка данных включает приведение размера изображений к кратному 32 (требование трансформера SegFormer) и стандартную нормализацию. На рисунке 2 приведена карта глубины, предсказанная с помощью данной модели.

Заключительным этапом проводится построение 3D-модели по предсказанной карте глубины. Постобработка карты глубины включает:

— удаление граничных артефактов (обрезка 16 пикселей);
— масштабирование (умножение на 1000, перевод в миллиметры);
— билатеральную фильтрацию для сглаживания с сохранением границ;
— медианную фильтрацию для удаления одиночных выбросов.

Построение 3D-модели выполняется с помощью библиотеки Open3D. Обратная проекция преобразует карту глубины в облако точек по формулам,

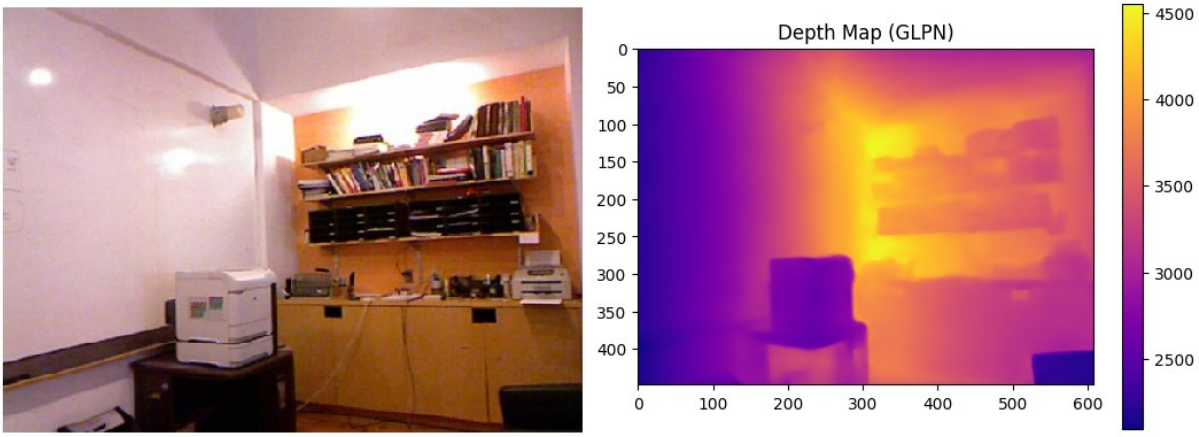

Рисунок 2 – Предсказанная карты глубины с помощью MiDaS

описанным ранее. Для построения полигональной сетки используется метод Пуассона. Результаты реконструкции представлены на рисунке 3.

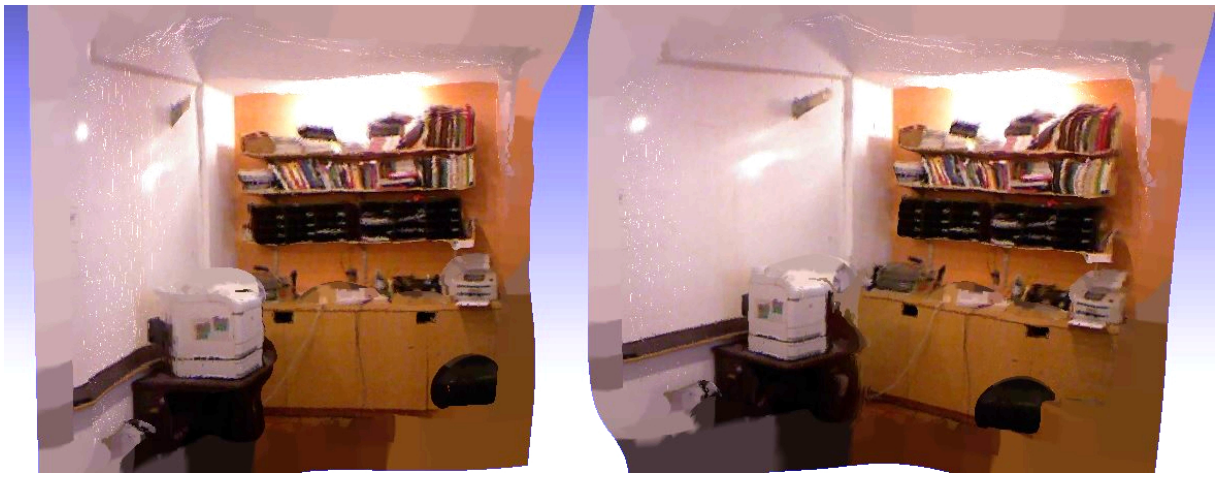

Рисунок 3 – Предсказанная карты глубины с помощью MiDaS

В подразделе 4.3 описан разработанный программный комплекс, реализованный на языке Python в среде Google Colab. Комплекс объединяет в себе проведение этапов, описанных ранее в данном разделе, и имеет пользовательский интерфейс, который представлен в виде консольного меню с пунктами:

1. Скачать NYU Dataset
2. Показать примеры NYU
3. Выбрать изображение из NYU
4. Загрузить своё изображение
5. Обучить MiDaS
6. Построить depth через MiDaS

7. Построить depth через GLPN
8. Построить 3D модель
9. Выход

В данном программном комплексе реализованы проверки наличия изображения перед построением карты глубины, проверка обученности модели, а также подключена возможность загружать свое изображение для построения его 3D-модели.

ЗАКЛЮЧЕНИЕ

Преобразование изображений из 2D в 3D является важным и перспективным направлением в области цифровых технологий. Этот процесс позволяет значительно сократить затраты ресурсов и времени, что особенно актуально в условиях современных требований к скорости и эффективности производства визуального контента. Автоматизированные методы преобразования позволяют создавать высококачественные трехмерные модели на основе обычных двумерных изображений, что открывает новые возможности для различных отраслей. Но в то же время не исключается важность специалистов по графческому дизайну, которые смогут доработать созданные автоматически объекты, привести их в наиболее рабочий вид.

Среди преимуществ данного подхода можно отметить упрощение реализации и экономию времени. В отличие от ручного создания 3D-моделей, которое требует значительных усилий и высокой квалификации, автоматическое преобразование позволяет быстро и эффективно получать качественные результаты. Помимо этого, воспользоваться нейросетевыми методами или готовой программой со встроенным искусственным интеллектом может даже новичок, что позволит создать конкурентоспособный продукт без больших вложений.

Помимо экономии времени, важными аспектом является возможность передать объем и детализированность сцен. Современные технологии с возможностями LIDAR или RGB-D позволяют создавать модели более правдоподобными и привлекательными для зрителей, сохранять физические размеры и пропорции. Это особенно важно в таких областях, как виртуальная и дополненная реальность, где качество визуального контента играет ключевую роль в создании погружающего и интерактивного опыта.

Таким образом, при написании выпускной квалификационной работы поставленная цель была достигнута, посредством выполнения всех необходимых для этого задач: изучены различные методы преобразования изображений в формате 2D в формат 3D, рассмотрены существующие программные и нейросетевые средства для создания 3D объектов из 2D изображений, создан комплекс для выполнения данной задачи на основе построения карты глубины, и проведено его тестирование.

Основные источники информации:

1 Войтешонок, Ю. В. Построение картины глубины изображения для портативных устройств / Ю. В. Войтешонок // Вестник Балтийского федерального университета им. И. Канта. Сер. Физико-математические и технические науки. 2022. №1. с. 14-20.

2 Кирпичников, А. П. Трёхмерная реконструкция сцены по нескольким изображениям / А. П. Кирпичников, И. И. Шамсутдинов, М. П. Шлеймович // Вестник Казанского технологического университета, vol. 17, no. 11, 2014. — с. 229-232.

3 Антонов, А. Сканирующие лазерные дальномеры (LIDAR) / А. Антонов // Современная электроника. — № 1, 2016. — С. 10-15.

4 Черноок, А. Ю. Применение RGB-D технологий в системах технического зрения / А. Ю. Черноок, Д. Ю. Перцев // 61-я научная конференция аспирантов, магистрантов и студентов БГУИР. — Минск, 2025. — С. 37-43.

5 Гафаров, Ф. М. Искусственные нейронные сети и приложения: учеб. пособие / Ф. М. Гафаров, А. Ф. Галимянов. — Казань: Изд-во Казан. ун-та, 2018. — 121 с.

6 Ranftl, R. Towards Robust Monocular Depth Estimation: Mixing Datasets for Zero-shot Cross-dataset Transfer / R. Ranftl, K. Lasinger, D. Hafner, K. Schindler, V. Koltun // IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). — 2020.

7 Kim, D. Global-Local Path Networks for Monocular Depth Estimation with Vertical CutDepth / D. Kim, W. Ga, P. Ahn, D. Joo, S. Chun, J. Kim // arXiv.org. — 2022.