УДК 004.8

Шамшура А.А.
Бакалавр
Головкина В.Б.
доцент
кафедры «автоматизированного
проектирования и дизайна»
Национальный исследовательский
технологический университет «МИСИС»

О ВОЗМОЖНОСТИ СОЗДАНИЯ КАРТ ГЛУБИНЫ ДЛЯ ВИДЕОКОНТЕНТА С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

Аннотация: Современную киноиндустрию и рекламное производство трудно представить без применения технологий цифровых визуальных эффектов (Visual Effects, VFX), ключевым элементом которых является композитинг. Для создания визуального оформления необходимо иметь техническую информацию, включающую карту глубины, получение которой требует больших финансовых затрат и профессионального оборудования. В данной работе экспериментальным путем обосновывается возможность применения нейросетевых технологий для генерации данных о глубине кадра для получения качественного изображения без существенных трудозатрат.

Ключевые слова: карта глубины, композитинг, нодовая система, нейросетевые технологии, comfyui, depth map, aov

Shamshura A.A.
Bachelor
Golovkina V.B.
Associate Professor
Department of Automated Design and Engineering
National University of Science and Technology «MISIS»

ON THE POSSIBILITY OF CREATING DEPTH MAPS FOR VIDEO CONTENT USING NEURAL NETWORK TECHNOLOGIES

Abstract: It is difficult to imagine the modern film industry and advertising production without the use of digital visual effects technologies (Visual Effects, VFX), the key element of which is compositing. To create a visual design, it is necessary to have technical information, including a depth map, which requires significant financial resources and professional equipment. This paper experimentally demonstrates the possibility of using neural network technologies to generate depth data for a frame in order to obtain a high-quality image without significant labor costs.

Keywords: depth map, compositing, node system, neural network technologies, comfyui, depth map, aov

Задача композитинга в визуальных эффектах заключается в корректной интеграции различных визуальных элементов в единое изображение. При работе со стереоскопическими материалами или при синтезе новых видов пространства для 3D-контента значительную роль играет информация о глубине сцены. Карта глубины (depth map) используется для проекции объектов между видами, управления порядком окклюзий и корректного размещения элементов в трёхмерном пространстве. Однако даже при наличии карт глубины процесс

композитинга остаётся технически сложным из-за ошибок, присутствующих в самих глубинных данных.

Как отмечают авторы публикации [1], композитинг стереоскопических сцен требует знания геометрии сцены в виде глубинной карты. Однако, получаемые традиционными методами, они являются недостаточно точными и содержат ошибки, которые негативно влияют на качество последующей обработки изображения. Для корректировки полученной информации ведется поиск новых методов, устойчивых к возникающим ошибкам, включая суперпиксельные алгоритмы, которые уменьшают чувствительность композитинга к шуму и неточностям глубинных данных. [2]

Таким образом, ограниченность существующих подходов к получению глубинных карт и возможные ошибки представляют собой фундаментальную проблему при выполнении композитинга в задачах компьютерной графики и пост-продакшена. Эта проблема особенно актуальна для сценариев, где исходная видеоинформация не содержит достоверных глубинных данных, что требует иных методов и подходов к созданию и корректировке карт глубины для различных видеоматериалов [3].

Оценка глубины изображения — фундаментальная задача компьютерного зрения, направленная на восстановление пространственной структуры сцены по двумерным данным [4]. Карта глубины кодирует расстояние от камеры до объектов сцены для каждого пикселя, что позволяет получить представление о трёхмерной геометрии сцены на основе изображений. В традиционных подходах вопрос решается с помощью стереовидения или специализированных сенсоров (LiDAR, RGB-D камеры), однако эти методы ограничены доступностью оборудования и условиями съёмки [5].

В последние годы появился метод монокулярной оценки глубины, позволяющий вычислить расстояние до объектов на сцене по одному изображению. Данная задача является качественно более сложной из-за «неопределённости» глубины по одиночному кадру, но современные методы глубокого обучения демонстрируют существенный прогресс в её решении, обеспечивая плотные карты глубины в end-to-end формате [6].

Ранние работы, касающиеся глубинной реконструкции, по одиночному изображению, использовали сверточные нейронные сети для предсказания глубины на основе многомасштабных признаков, обучаемые на наборе данных с известной глубиной. Эти подходы показали, что нейросети способны учитывать сложные визуальные признаки перспективы и текстуры для получения правдоподобных карт глубины [7].

Помимо монокулярных методов, значительное внимание уделяется стерео-глубине, где входными данными служат пара изображений, а глубина восстанавливается на основе сопоставления признаков в пространстве диспаратности. Такие методы, включая глубокие сети с cost-volume и siamese-архитектурами, демонстрируют более высокую точность, но требуют наличия двух согласованных по калибровке изображений [8].

В настоящее время существует ряд нейросетевых моделей и программных решений, применяемых для восстановления карт глубины изображения и используемых в VFX-пайплайнах, в том числе на этапе композитинга. Информация о глубине позволяет упростить интеграцию визуальных эффектов, имитацию глубины резкости и атмосферной перспективы.

В коммерческой среде распространены решения, интегрируемые в профессиональные пакеты постпродакшена, такие как Adobe After Effects, включая плагины Depth Scanner, Volumax и инструменты KeenTools. Эти продукты ориентированы на профессиональное использование и поддерживают работу с видеоматериалом, однако являются платными [9].

Среди открытых нейросетевых моделей наиболее широко используется MiDaS, а её развитием является модель DPT (Depth Prediction Transformer), основанная на архитектуре трансформеров и обеспечивающая более высокую точность на сложных сценах. Вместе с тем данные модели ориентированы на обработку одиночных изображений и не обеспечивают временную согласованность при работе с видео, что ограничивает их прямое применение в промышленном композитинге. [5]

Возможным решением становится использование Comfy UI в сочетании с нейросетью depth_anything_ver2. Это программа с открытым исходным кодом, позволяющая пользователям генерировать изображения на основе текстовых запросов. В Comfy UI есть библиотеки дополнений, которые поддерживают импорт и экспорт видео, а также секвенций (массивов) изображений.

Для выполнения эксперимента собрана минимальная нодовая система, предназначенная для генерации карты глубины на основе видеопоследовательности. Конфигурация графа является относительно простой и включает три основных ноды, однако корректность результата напрямую зависит от точности исходных настроек. К критически важным параметрам относятся формат экспорта, интерпретация видеоматериала, его пространственное разрешение, цветовое пространство и частота кадров (FPS) (рис.1.)

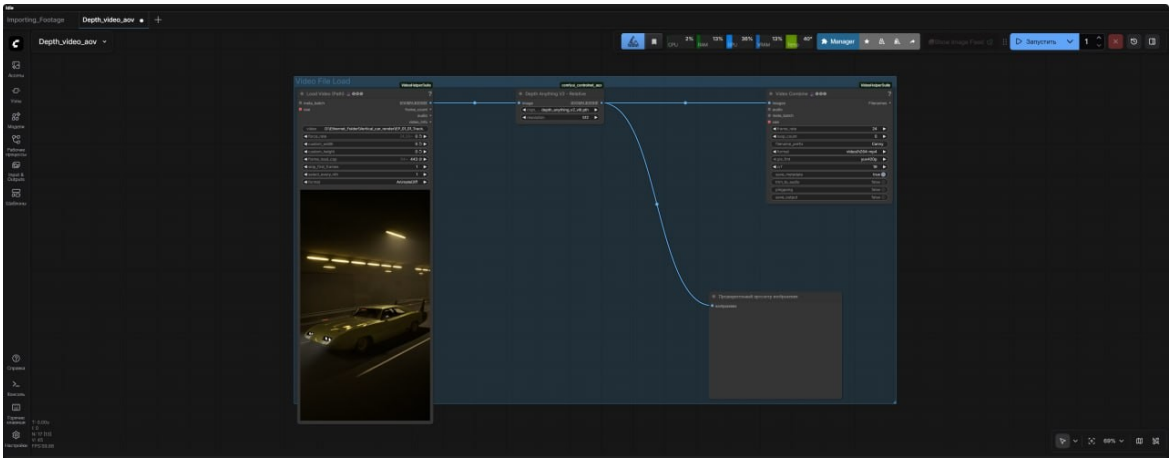


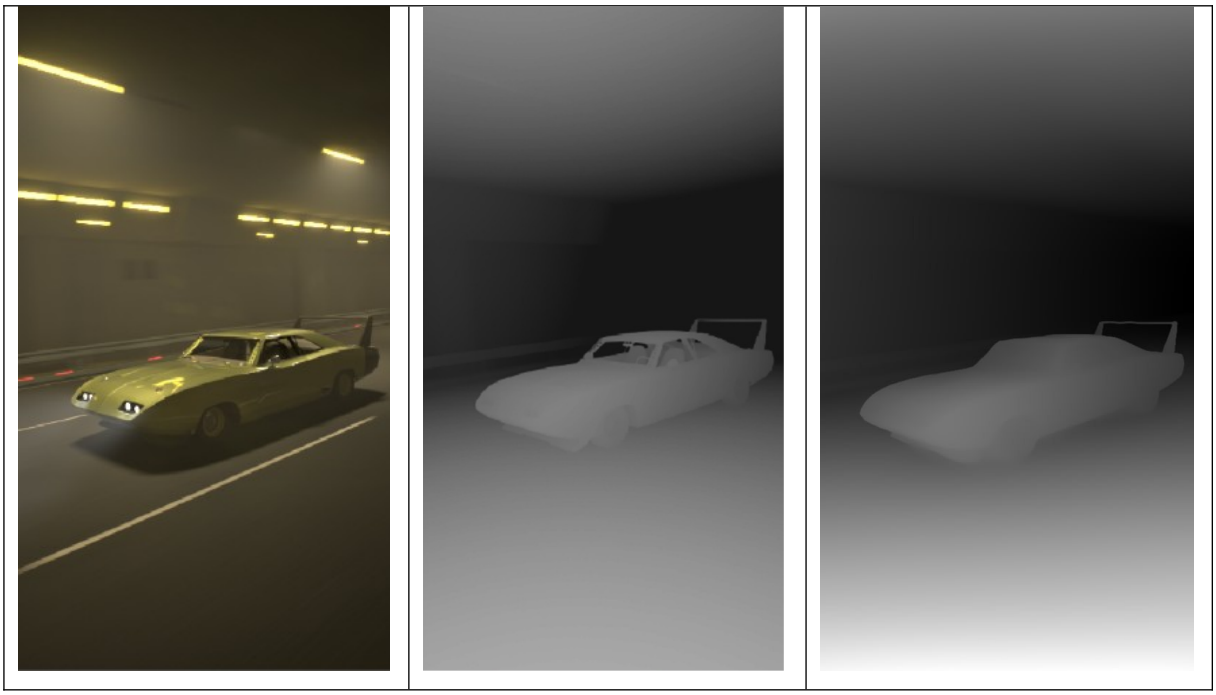
Рис. 1. Пример нодового графа для генерации изображения

В качестве входных данных используется один из тестовых рендеров, ранее созданных в игровом движке Unreal Engine. Данный видеоматериал рассматривается как имитация съемки с реальной камеры и применяется для последующего сравнения нейросетевой карты глубины и карты глубины, полученной непосредственно в Unreal Engine.

После завершения настройки вычислительного графа запущен процесс генерации. Расчёт карты глубины выполнялся на одной видеокарте NVIDIA RTX 3090 и занял около 20 секунд. Указанное время актуально для видеопоследовательности длиной 24 кадра с разрешением 2K (2560×1440).

Таблица 1 – Сравнение полученных пассов глубины

Обычный пасс	Unrel Engine пасс глубины	AI пасс глубины



Сравнительный анализ представленных карт глубины показывает, что нейросетевая модель корректно интерпретирует автомобиль как цельный объемный объект. В то же время карта гдубины, полученная в Unreal Engine, демонстрирует некорректную обработку прозрачных элементов: стекла автомобиля отображают глубину объектов, расположенных за ними. Данное поведение обусловлено техническими ограничениями рендеринга глубины в Unreal Engine и типично для Z-buffer–основанных решений при работе с прозрачными материалами. Нельзя не заметить, что у нейросетевой карты глубины куда менее четкие границы объектов, что свидетельствует о не достаточно высокой точности при работе с данной нейросетью, но полученный результат подходит для дальнейшего применения в видеопроизводстве.

Консистентность (постоянство) генерации подтвержено при следующем использовании, а анализ полученых данных с каждого кадра выявил 10% отклонение в значениях по глубине кадра между кадрами одного видео. Это отклоение не критично для добавление размытия на задний план или создания расфокусировки на переднем.

В рамках данной работы экспериментально подтверждена практическая применимость нейросетевых технологий для генерации карт глубины видеоконтента в задачах композитинга и VFX. Проведённый анализ показал, что использование современных моделей монокулярной оценки глубины позволяет получать информативные и визуально согласованные depth map без применения специализированного оборудования и дорогостоящих сенсоров, что существенно снижает порог входа в производственный процесс.

Эксперимент с использованием связки Comfy UI и нейросетевой модели *depth_anything_ver2* продемонстрировал, что даже при минимальной нодовой конфигурации возможно получение качественных карт глубины, пригодных для последующего использования в композитинге. Сравнение нейросетевой карты глубины с пассом глубины, полученным в Unreal Engine, выявило ряд преимуществ нейросетевого подхода, в частности более корректную интерпретацию сложных объектов и прозрачных материалов. Это значит, что данная технология применима для генерации карт глубин для видеоматериалов, снятых на камеру, где получение информации о глубине кадра либо невозможно, либо связано с большими финансовыми затратами на оборудование. Это важно для задач постпродакшена, где ошибки глубины напрямую влияют на корректность окклюзий, глубины резкости и атмосферных эффектов.

Таким образом, нейросетевая генерация карт глубины является перспективным направлением развития инструментов компьютерной графики и цифрового кинопроизводства.

Список литературы

1. Schnyder L. et al. Depth image based compositing for stereo 3D //2012 3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON). – IEEE, 2012. – С. 1-4.
2. DEPTHIMAGEBASEDCOMPOSITINGFORSTEREO 3D [Электронный ресурс]. URL: https://la.disneyresearch.com/wp-content/uploads/Depth-Image-Based-Compositing-for-Stereo-3D-Paper.pdf (дата обращения 15.12.2025)
3. Есаков А. А., Дуплей М. И. Применение искусственного интеллекта в анимации и кинопроизводстве //Вестник науки. – 2025. – Т. 1. – №. 12 (93). – С. 1929-1936.
4. Морев К. И., Ледерер П. А. Экспериментальная оценка погрешностей восстановления структуры наблюдаемой сцены из серии снимков движущейся камеры //Известия Южного федерального университета. Технические науки. – 2024. – №. 1 (237). – С. 276-285.
5. Истомин В. И., Привалов А. Н. К вопросу применимости методов монокулярной оценки глубины для 3D-реконструкции геометрии документов //Известия Тульского государственного университета. Технические науки. – 2024. – №. 9. – С. 401-405.
6. Monocular Depth Estimation Based On Deep Learning: An Overview [Электронный ресурс]. URL: https://arxiv.org/abs/2003.06620 (дата обращения 20.12.2025)
7. Single Image Depth Estimation: An Overview [Электронный ресурс]. URL: https://arxiv.org/abs/2104.06456 (дата обращения 22.12.2025)
8. Зуйков И. В. Особенности использования искусственного интеллекта в кинематографе и медиаиндустрии //Вестник ВГИК. – 2022. – Т. 14. – №. 4 (54). – С. 65-77.

9. Евич Л. Н., Калинин П. А., Еременко С. А. Плагины, как инструмент реализации новых возможностей //Российская наука в современном мире. – 2017. – С. 91-92.