



Международный журнал информационных технологий и  
энергоэффективности

Сайт журнала:

<http://www.openaccessscience.ru/index.php/ijcse/>



УДК 004.032.26

## СИНТЕЗ НОВЫХ ВИДОВ СЛОЖНОЙ СЦЕНЫ ПО РЕАЛЬНЫМ ФОТОГРАФИЯМ С ПОМОЩЬЮ NERF

**Чернов Е.И.**

*Московский Политехнический Университет, Россия (107023, Москва, Большая Семеновская ул., 38), e-mail: chernov.eu.i@gmail.com*

Синтез новых видов сцены давняя проблема в области компьютерного зрения. На данный момент существует множество подходов к решению названной задачи, от обычной интерполяции между отдельными видами до нейронного рендеринга. В данной работе рассматриваются синтез новых видов сложной сцены по реальным фотографиям. В точности для исследования используются фотографии здания в жилом квартале снятое с разных ракурсов с различными изменчивыми объектами такими, как движущиеся машины или пешеходы, является значительной помехой для получения качественного результата. Благодаря NeRF (Neural Radiance Fields) был совершён прорыв. В контексте рассматриваемой проблематики можно выделить NeRF-W, позволяющего работать с фотографиями, сделанными при разных условиях освещения и наличии изменчивых объектов, и Instant NeRF, работающий в разы быстрее аналогов.

Ключевые слова: синтез видов, рендеринг на основе изображений, объемный рендеринг, 3d глубокое обучение.

## SYNTHESIS OF NEW VIEWS OF COMPLEX SCENES FROM REAL PHOTOS USING NERF

**Chernov E.I.**

*Moscow Polytechnic University, Russia (107023, Moscow, Bolshaya Semenovskaya st., 38), e-mail: chernov.eu.i@gmail.com*

Synthesis of views of scenes is a long-standing problem in the field of computer vision. At the moment, there are many approaches to solving this problem, from the usual interpolation between individual views to neural rendering. This paper considers the synthesis of views of complex scenes from real photographs. Exactly for the research, photographs of a building in a residential quarter are used, taken from different angles with various changeable objects such as moving cars or pedestrians, is a significant obstacle to obtaining a high-quality result. Thanks to NeRF (Natural Radiance Fields), a breakthrough was made. In the context of the issues under consideration, we can distinguish NeRF-W, which allows you to work with photos taken under different lighting conditions and the presence of changeable objects, and Instant NeRF, which works many times faster than analogues.

Keywords: scene representation, view synthesis, image-based rendering, volume rendering, 3D deep learning.

### Введение

Синтез новых видов сцены давняя проблема в области компьютерного зрения. На данный момент существует множество подходов к решению названной задачи, от обычной интерполяции между отдельными видами до нейронного рендеринга. Современные методы

способны синтезировать новые виды с удивительно точной детализацией, хотя построение 3D моделей таких, как полигональная сетка остается сложной для решения задачей.

В данной работе рассматриваются синтез новых видов сложной сцены по реальным фотографиям. В точности для исследования используются фотографии здания в жилом квартале снятое с разных ракурсов с различными изменчивыми объектами такими, как движущиеся машины или пешеходы, является значительной помехой для получения качественного результата.

В области синтеза новых видов сложной сцены был совершён прорыв благодаря NeRF (Neural Radiance Fields) [1]. Данный метод предлагает использовать многослойный перцептрон для сопоставления каждой точке пространства, объемной плотности и цвета. На момент написания данной работы существует уже множество разновидностей NeRF, улучшающие метод по тем или иным показателям. В контексте рассматриваемой проблематики можно выделить NeRF-W [2], позволяющего работать с фотографиями, сделанными при разных условиях освещения и наличии изменчивых объектов, и Instant NeRF [3], работающий в разы быстрее аналогов.

В этом исследовании производится поиск наилучшего существующего решения для синтеза новых видов сложных сцен.

Решение названной проблемы позволит получать сложные объекты реального мира в цифровом формате для последующего анализа или встраивания в виртуальное пространство.

### **Предыдущие работы.**

Синтез новых видов является давней проблемой в области компьютерного зрения. На основе плотной выборки видов могут быть использованным классические методы интерполяции [4–6].

В последнее время были достигнуты значительные достижения в синтезе видов из разреженного набора данных. Один популярный класс подходов использует представления сцен на основе полигональной сетки с размытым [7] либо зависящим от вида [8; 9] изображением.

Дифференцируемые растризаторы [10–13] или трассировщики путей [14] могут напрямую оптимизировать представления полигональной сетки для воспроизведения набора входных изображений с использованием градиентного спуска. Однако оптимизация сетки на основе градиента, основанная на перепроекции изображения, часто затруднена, вероятно, из-за локальных минимумов или плохой подготовки ландшафта потерь. Кроме того, эта стратегия требует, чтобы шаблонная полигональная сетка с фиксированной топологией предоставлялась в качестве инициализации перед оптимизацией [14], что обычно недоступно для неограниченных сцен реального мира.

Другой класс методов использует объемные представления для решения задачи синтеза высококачественного фотореалистичного вида из набора входных изображений RGB. Объемные подходы способны реалистично представлять сложные формы и материалы, хорошо подходят для оптимизации на основе градиентов и, как правило, создают меньше визуально отвлекающих артефактов, чем методы на основе полигональной сетки. Ранние объемные подходы использовали наблюдаемые изображения для непосредственного окрашивания воксельных сеток [15–17]. Несколько методов [18–25] использовали большие наборы данных из нескольких сцен для обучения глубоких сетей, которые предсказывают

выборочное объемное представление из набора входных изображений, а затем используют либо альфа-композицию [26], либо научное композирование вдоль лучей для визуализации новых видов во время тестирования.

В других работах была оптимизирована комбинация сверточных сетей (CNN) и выборочных воксельных сеток для каждой конкретной сцены, так что CNN может компенсировать артефакты дискретизации от воксельных сеток низкого разрешения [27] или позволить прогнозируемому сетки вокселей могут изменяться в зависимости от времени ввода или элементов управления анимацией [28]. В то время как эти объемные методы достигли впечатляющих результатов для синтеза новых изображений, их способность масштабироваться до изображений с более высоким разрешением принципиально ограничена низкой сложностью во времени и пространстве из-за их дискретной выборки — для рендеринга изображений с более высоким разрешением требуется более точная выборка 3D-пространства.

Структура из движения [29] и настройка связки [30] могут быть использованы для построения разряженного облака точек и восстановления параметров камеры. На примере набора данных Photo Tourism [31], было показано, что эти методы могут быть масштабированы до огромных наборов и использованы для синтеза новых видов [32; 33].

Недавние работы явно выводят свойства света и отражения объектов в сцене из набора неограниченных коллекций фотографий [34–36], используя их для управления внешним видом сцены и геометрией. В то время как другие методы используют семантические знания для реконструкции переходных объектов [37].

### **Нейронный рендеринг.**

Методы нейронного рендеринга начали применяться с недавних пор [38]. В нескольких подходах используются сети трансляции изображений [39] для более реалистичного отображения контента, используя в качестве входных данных традиционные результаты реконструкции [40], изученные скрытые текстуры [41], облака точек [42], воксели [27] или объемы развертки плоскости [18; 43]. Общими недостатками этих подходов являются временные артефакты, видимые при движении камеры.

Другой недавний подход представляет сцену в виде многоплоскостных изображений, ориентированных на камеру, для реконструкции захваченных сцен [21; 25] и интернет-коллекций фотографий [44]. Эти методы позволяют получить фотореалистичную визуализацию с новых точек зрения, но виды, которые они могут интерполировать, ограничены. Напротив, подходы к объемному рендерингу [28; 45] позволяют получать точные и последовательные реконструкции даже при больших движениях камеры, как это делает NeRF-W [2]. Neural Radiance Fields (NeRF) [1] используют многослойный перцептрон (MLP) для моделирования поля излучения с беспрецедентным уровнем точности, отчасти благодаря использованию позиционного кодирования в MLP [46]. Однако оптимизация NeRF может быть довольно продолжительной. Учитывая то, что NeRF оптимизируется для каждой сцены отдельно, это может быть серьезным затруднением. Для преодоления названного затруднения могут применяться разные инструменты, включая Instant NeRF, который позволяет значительно сократить время, затраченное на оптимизацию [3].

В данной работе будут рассмотрены классический NeRF, Instant NeRF и NeRF-W.

### Neural Radiance Fields (NeRF)

NeRF — это метод для синтеза новых видов сцены, представленной в виде излученного непрерывного объемного поля. В основе его лежит многослойный персептрон (MLP). В качестве входа выступают декартовы координаты точки на сцене  $\mathbf{x} = (x, y, z)$  и направление просмотра  $\mathbf{d} = (d_x, d_y, d_x)$ , а в качестве выхода — объемная плотность  $\sigma$  и цвет  $\mathbf{c} = (r, g, b)$ . Чтобы вычислить цвет одного пикселя, аппроксимируется интеграл объемного рендеринга, используя числовую квадратуру [47]. Пусть  $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$  — луч, испускаемый из центра проекции камеры  $\mathbf{o}$  через данный пиксель на плоскости изображения, тогда цвет этого пикселя будет вычисляться в соответствии с выражением 1.

$$\hat{\mathbf{c}}(\mathbf{r}) = \sum_{k=1}^K T(t_k) \alpha(\sigma(t_k) \delta_k) \mathbf{c}(t_k),$$

$$T(t_k) = \exp\left(-\sum_{n=1}^{k-1} \sigma(t_n) \delta_n\right),$$

$$\alpha(x) = 1 - \exp(-x),$$
(1)

где  $\hat{\mathbf{c}}$  — вычисляемый цвет пикселя,  $\{t_k\}_{k=1}^K$  — квадратурные точки на луче  $\mathbf{r}$  между ближней и дальней плоскостями камеры,  $\sigma(t)$  и  $\mathbf{c}(t)$  — объемная плотность и цвет в точке  $\mathbf{r}(t)$ , и  $\delta_k = t_{k+1} - t_k$  расстояние между двумя последовательными квадратурными точками.

Функция потерь считается на основании действительного изображения и изображения полученного с помощью выражения 1. Архитектура модели представлена на Рисунке 1.

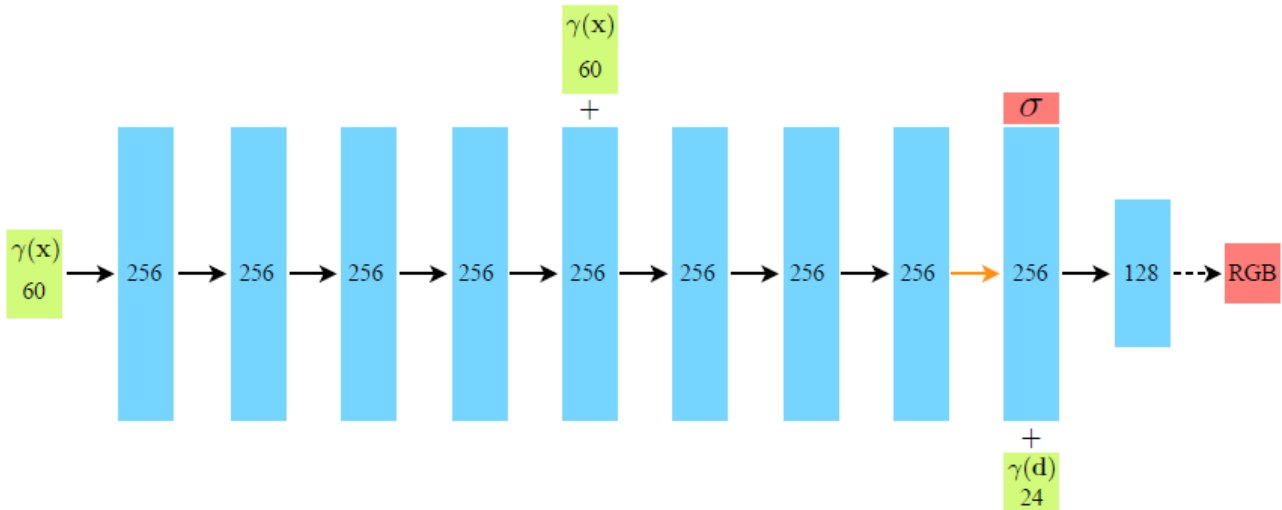


Рисунок 1 – Архитектура NeRF [1]

При этом данный метод имеет несколько особенностей, позволяющих ему достигать конкурентного качества:

- 1) модель регуляризируется благодаря разделению выходов объемной плотности и цвета на разные слои;
- 2) чтобы увеличить качество высокочастотных деталей сцены, к декартовым координатам и направлению просмотра применяется позиционное кодирование  $\gamma$ ;
- 3) используется грубая и мелкая модели, которые обучаются друг за другом, позволяя увеличивать качество результата в значимых областях пространства.

### **NeRF in the Wild (NeRF-W)**

NeRF-W развивает идеи NeRF, распространяясь на фотографии, сделанные в разных условиях. В названном подходе решаются две проблемы:

- фотометрическая изменчивость: фотографии, сделанные при разных условиях освещения;
- временные объекты: в объектив камеры могут попадать случайные объекты отсутствующие на других фотографиях и движущиеся объекты.

Данные пункты влияют на итоговую реконструкцию сцены, добавляя шум или искажая цвета. Для решения названных проблем применяется генеративная латентная оптимизация (Generative Latent Optimization – GLO) [48]. Данный подход состоит в том, чтобы описывать каждый обучающий пример своим вектором в латентном пространстве. При этом эти вектора подаются на вход модели и оптимизируются вместе с ее весами.

Для решения проблемы фотометрической изменчивости, каждому изображению сопоставляется свой вектор, описывающий условия освещенности. Также добавляется дополнительный вывод цвета, соответствующего условиям освещенности переданного вектора.

Проблема временных объектов решается похожим образом. Каждой фотографии сопоставляется еще по вектору для описания временных объектов. И добавляются дополнительные выходы цвета, объемной плотности и неопределенности. Неопределенность описывает с какой вероятностью пиксель того или иного изображения относился к временному объекту.

На Рисунке 2 представлена архитектура NeRF-W.

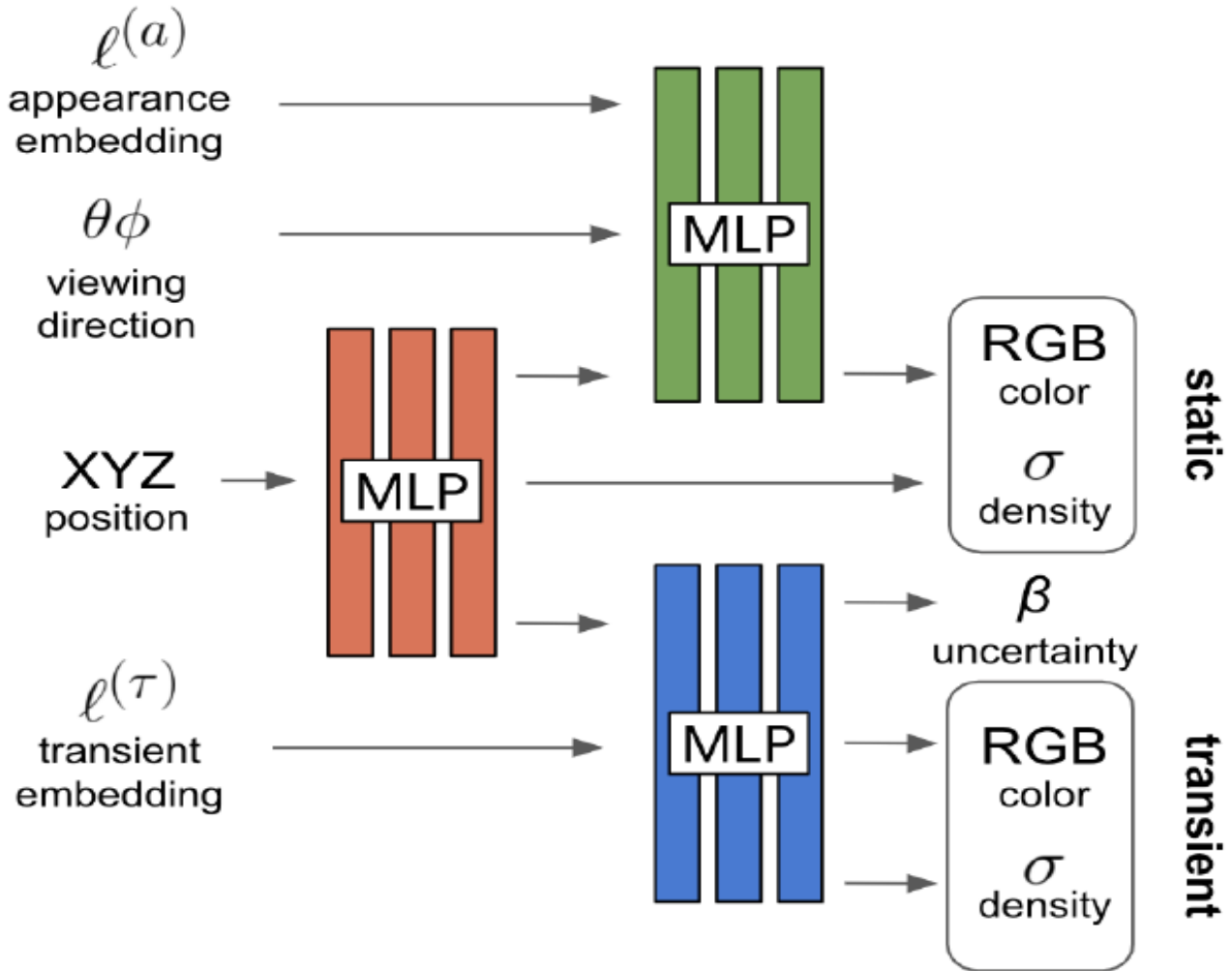


Рисунок 2 – Архитектура NeRF-W [2]

### Instant NeRF

Instant NeRF использует архитектуру схожую с классическим NeRF, но заменяет позиционное кодирование хеш-кодированием с мультиразрешением. Это позволяет упростить архитектуру NeRF, снижая количество операций с плавающей точкой.

### Эксперименты.

Для проведения экспериментов было отснято некоторое здание с нескольких ракурсов. При этом в кадр попадает несколько изменчивых объектов, таких как движущиеся машины или пешеходы. Несколько ракурсов приведены на Рисунке 3.



Рисунок 3 – Примеры из обучающего набора данных

Для оптимизации параметров каждого метода использовалось 185 изображений здания с разных ракурсов с разрешением 3840x2160. Для оценки параметров камеры использовался COLMA [49] с включением двух радиальных и двух тангенциальных параметров. Модель инициализируется случайными весами. Для оптимизации каждого варианта NeRF использовался Adam на 1 GPU Tesla A100 с размером пакета 34816 на протяжении 20 эпох. Это потребовало около 6 часов для классического NeRF, около 10 часов для NeRF-W и несколько десятков минут для Instant NeRF.

Так как сравнение качества синтеза новых видов является сложной задачей, на Рисунке 4 приводятся результаты каждого метода с одного и того же ракурса. Количественное сравнение качества на основе PSNR, MS-SSIM [50] и LPIPS [51] приведено в Таблице 1.

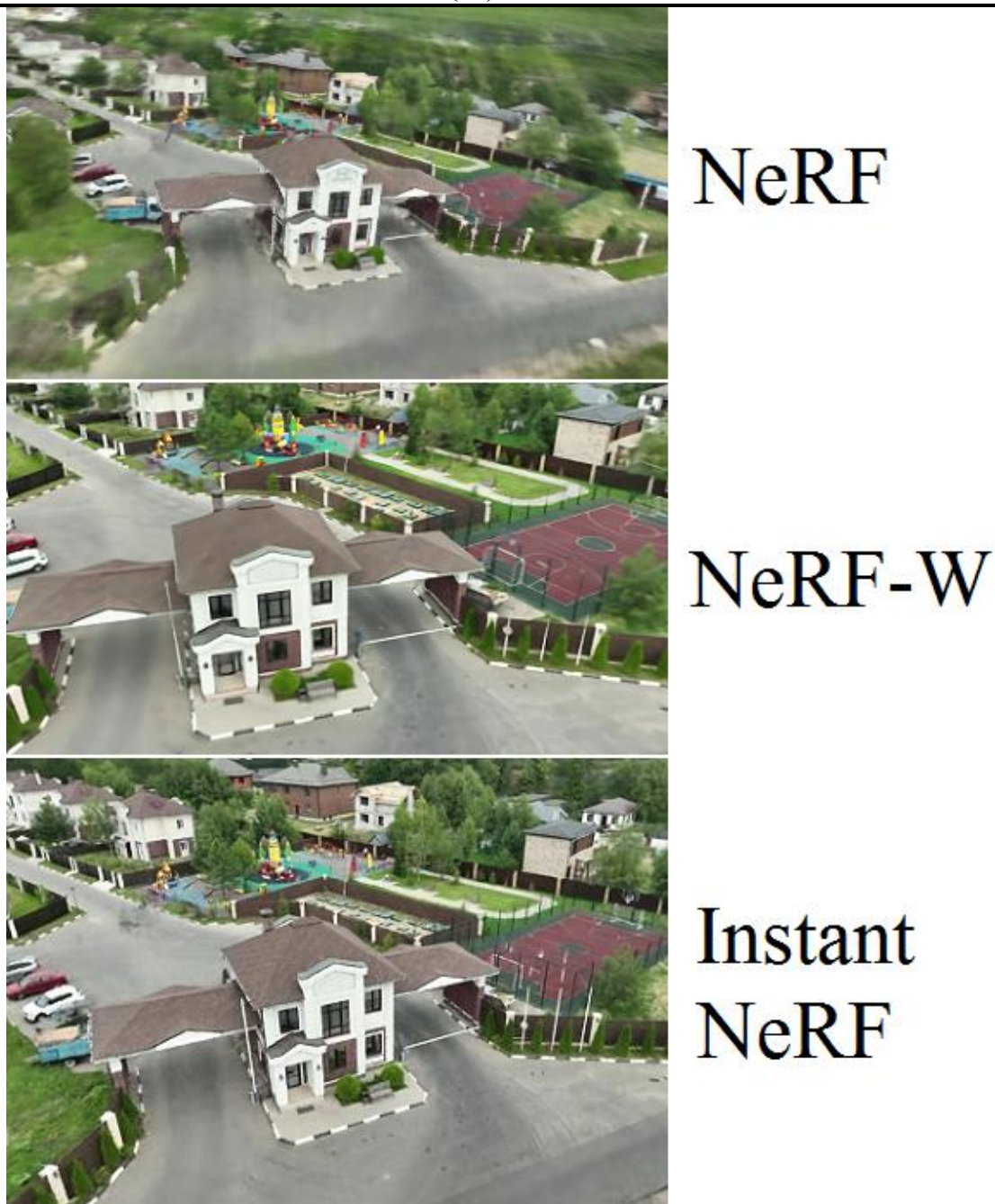


Рисунок 4 – Результаты использования каждого из методов

Таблица 1 – Количественное сравнение качества на основе PSNR, MS-SSIM и LPIPS. В таблице выделены наилучшие метрики.

	PSNR	MS-SSIM	LPIPS
NeRF	22,19	0,895	0,192
NeRF-W	<b>25,32</b>	<b>0,946</b>	<b>0,133</b>
Instant NeRF	23,03	0,903	0,175

На основании полученных результатов видно, что NeRF-W обладает наилучшими результатами. Это объясняется наличием множества изменчивых объектов на изображениях исследуемого набора данных.



### **Заключение.**

В данной работе был проведен ряд экспериментов по синтезу новых видов сложных сцен по реальным фотографиям с помощью современных методов по нейронному рендерингу NeRF, NeRF-W, Instant NeRF. Результаты показывают, что NeRF-W значительно превосходит по качеству другие методы на рассматриваемом наборе данных, однако требует значительные объемы времени на оптимизацию.

Синтез новых видов рассмотренной сцены был выполнен на приемлемом уровне. Для получения наиболее качественных результатов на схожих наборах данных наиболее подходящим является NeRF-W. Для получения быстрых результатов с достойным качеством может использоваться Instant NeRF.

### **Список литературы**

1. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. NeRF / В. Mildenhall [и др.] number: arXiv:2003.08934arXiv:2003.08934 [cs]. – arXiv, 2020.
2. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. NeRF in the Wild / R. Martin-Brualla [и др.] number: arXiv:2008.02268arXiv:2008.02268 [cs]. – arXiv, 2021.
3. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding / Т. Müller [и др.] // ACM Transactions on Graphics. – 2022. – Т. 41. – № 4. – С. 1-15.
4. Levoy M. Light field rendering / М. Levoy, Р. Hanrahan // Proceedings of the 23rd annual conference on Computer graphics and interactive techniques : SIGGRAPH '96. – New York, NY, USA: Association for Computing Machinery, 1996. – С. 31-42.
5. Davis A. Unstructured Light Fields / А. Davis, М. Levoy, F. Durand // Computer Graphics Forum. – 2012. – Т. 31.
6. The Lumigraph / S. Gortler [и др.] // Proc. of SIGGRAPH 96. – 2001. – Т. 96.
7. Waechter M. Let There Be Color! Large-Scale Texturing of 3D Reconstructions / М. Waechter, N. Moehrle, M. Goesele // Computer Vision – ECCV 2014 : Lecture Notes in Computer Science / eds. D. Fleet [et al.]. – Cham: Springer International Publishing, 2014. – С. 836-850.
8. Unstructured lumigraph rendering / С. Buehler [и др.] // Proceedings of the 28th annual conference on Computer graphics and interactive techniques : SIGGRAPH '01. – New York, NY, USA: Association for Computing Machinery, 2001. – С. 425-432.
9. Surface light fields for 3D photography / D.N. Wood [и др.] // Proceedings of the 27th annual conference on Computer graphics and interactive techniques : SIGGRAPH '00. – USA: ACM Press/Addison-Wesley Publishing Co., 2000. – С. 287-296.
10. Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer / W. Chen [и др.] number: arXiv:1908.01210arXiv:1908.01210 [cs]. – arXiv, 2019.
11. Unsupervised Training for 3D Morphable Model Regression / К. Genova [и др.] number: arXiv:1806.06098arXiv:1806.06098 [cs]. – arXiv, 2018.
12. Soft Rasterizer: A Differentiable Renderer for Image-based 3D Reasoning. Soft Rasterizer / S. Liu [и др.] number: arXiv:1904.01786arXiv:1904.01786 [cs]. – arXiv, 2019.
13. Loper M.M. OpenDR: An Approximate Differentiable Renderer / М.М. Loper, М.М. Black // Computer Vision – ECCV 2014 : Lecture Notes in Computer Science / eds. D. Fleet [et al.]. – Cham: Springer International Publishing, 2014. – OpenDR. – С. 154-169.

14. Differentiable Monte Carlo ray tracing through edge sampling / T.-M. Li [и др.] // ACM Transactions on Graphics. – 2018. – Т. 37. – № 6. – С. 222:1-222:11.
15. Kutulakos K.N. A Theory of Shape by Space Carving / K.N. Kutulakos, S.M. Seitz // International Journal of Computer Vision. – 2000. – Vol.38. – № 3. – С. 199-218.
16. Seitz S.M. Photorealistic Scene Reconstruction by Voxel Coloring / S.M. Seitz, C.R. Dyer // International Journal of Computer Vision. – 1999. – Vol. 35. – № 2. – С. 151-173.
17. Szeliski R. Stereo Matching with Transparency and Matting / R. Szeliski, P. Golland // International Journal of Computer Vision. – 1999. – Vol. 32. – № 1. – С. 45-61.
18. DeepView: View Synthesis with Learned Gradient Descent. DeepView / J. Flynn [и др.] number: arXiv:1906.07316arXiv:1906.07316 [cs, eess]. – arXiv, 2019.
19. Single-image Tomography: 3D Volumes from 2D Cranial X-Rays. Single-image Tomography / P. Henzler [и др.] number: arXiv:1710.04867arXiv:1710.04867 [cs]. – arXiv, 2018.
20. Kar A. Learning a Multi-View Stereo Machine / A. Kar, C. Häne, J. Malik // Advances in Neural Information Processing Systems. – Curran Associates, Inc., 2017. – Т. 30.
21. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. Local Light Field Fusion / B. Mildenhall [и др.] number: arXiv:1905.00889arXiv:1905.00889 [cs]. – arXiv, 2019.
22. Penner E. Soft 3D reconstruction for view synthesis / E. Penner, L. Zhang // ACM Transactions on Graphics. – 2017. – Т. 36. – № 6. – С. 235:1-235:11.
23. Pushing the Boundaries of View Extrapolation with Multiplane Images / P.P. Srinivasan [и др.] number: arXiv:1905.00413arXiv:1905.00413 [cs]. – arXiv, 2019.
24. Multi-view Supervision for Single-view Reconstruction via Differentiable Ray Consistency / S. Tulsiani [и др.] number: arXiv:1704.06254arXiv:1704.06254 [cs]. – arXiv, 2017.
25. Stereo Magnification: Learning View Synthesis using Multiplane Images. Stereo Magnification / T. Zhou [и др.] number: arXiv:1805.09817arXiv:1805.09817 [cs]. – arXiv, 2018.
26. Porter T. Compositing digital images / T. Porter, T. Duff // ACM SIGGRAPH Computer Graphics. – 1984. – Vol. 18. – № 3. – С. 253-259.
27. DeepVoxels: Learning Persistent 3D Feature Embeddings. DeepVoxels / V. Sitzmann [и др.] number: arXiv:1812.01024arXiv:1812.01024 [cs]. – arXiv, 2019.
28. Neural Volumes: Learning Dynamic Renderable Volumes from Images / S. Lombardi [и др.] // ACM Transactions on Graphics. – 2019. – Т. 38. – Neural Volumes. – № 4. – С. 1-14.
29. Hartley R. Multiple View Geometry in Computer Vision / R. Hartley, A. Zisserman. – 2. – Cambridge: Cambridge University Press, 2004.
30. Bundle Adjustment — A Modern Synthesis / B. Triggs [et al.] // Vision Algorithms: Theory and Practice : Lecture Notes in Computer Science / eds. B. Triggs, A. Zisserman, R. Szeliski. – Berlin, Heidelberg: Springer, 2000. – С. 298-372.
31. Snavely N. Photo tourism: exploring photo collections in 3D / N. Snavely, S.M. Seitz, R. Szeliski // ACM Transactions on Graphics. – 2006. – Т. 25. – Photo tourism. – № 3. – С. 835-846.
32. Reconstructing Rome / S. Agarwal [и др.] // Computer. – 2010. – Т. 43. – № 6. – С. 40-47.
33. Building Rome on a Cloudless Day / J.-M. Frahm [et al.] // Computer Vision – ECCV 2010 : Lecture Notes in Computer Science / eds. K. Daniilidis, P. Maragos, N. Paragios. – Berlin, Heidelberg: Springer, 2010. – С. 368-381.

34. Coherent Intrinsic Images from Photo Collections : Proceedings of ACM SIGGRAPH Asia 2012 / P.-Y. Laffont [et al.] // ACM Transactions on Graphics. – 2012. – Vol. 31. – № 6.
35. The Visual Turing Test for Scene Reconstruction / Q. Shan [и др.] // 2013 International Conference on 3D Vision - 3DV 2013 2013 International Conference on 3D Vision - 3DV 2013. – 2013. – С. 25-32.
36. Learning to Factorize and Relight a City / A. Liu [и др.] // European Conference on Computer Vision (ECCV). – 2020.
37. Augmenting Crowd-Sourced 3D Reconstructions Using Semantic Detections / T. Price [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – С. 1926-1935.
38. State of the Art on Neural Rendering / A. Tewari [и др.] number: arXiv:2004.03805arXiv:2004.03805 [cs]. – arXiv, 2020.
39. Image-to-Image Translation with Conditional Adversarial Networks / P. Isola [и др.] number: arXiv:1611.07004arXiv:1611.07004 [cs]. – arXiv, 2018.
40. LookinGood: Enhancing Performance Capture with Real-time Neural Re-Rendering. LookinGood / R. Martin-Brualla [и др.] number: arXiv:1811.05029arXiv:1811.05029 [cs]. – arXiv, 2018.
41. Thies J. Deferred Neural Rendering: Image Synthesis using Neural Textures. Deferred Neural Rendering / J. Thies, M. Zollhöfer, M. Nießner number: arXiv:1904.12356arXiv:1904.12356 [cs]. – arXiv, 2019.
42. Neural Point-Based Graphics / K.-A. Aliev [и др.] number: arXiv:1906.08240arXiv:1906.08240 [cs]. – arXiv, 2020.
43. DeepStereo: Learning to Predict New Views from the World’s Imagery. DeepStereo / J. Flynn [и др.] number: arXiv:1506.06825arXiv:1506.06825 [cs]. – arXiv, 2015.
44. Crowdsampling the Plenoptic Function / Z. Li [и др.] number: arXiv:2007.15194arXiv:2007.15194 [cs]. – arXiv, 2020.
45. Sitzmann V. Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. Scene Representation Networks / V. Sitzmann, M. Zollhöfer, G. Wetzstein number: arXiv:1906.01618arXiv:1906.01618 [cs]. – arXiv, 2020.
46. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains / M. Tancik [и др.] number: arXiv:2006.10739arXiv:2006.10739 [cs]. – arXiv, 2020.
47. Max N. Optical models for direct volume rendering / N. Max // IEEE Transactions on Visualization and Computer Graphics. – 1995. – Т. 1. – № 2. – С. 99-108.
48. Optimizing the Latent Space of Generative Networks / P. Bojanowski [и др.] number: arXiv:1707.05776arXiv:1707.05776 [cs, stat]. – arXiv, 2019.
49. Schonberger J.L. Structure-From-Motion Revisited / J.L. Schonberger, J.-M. Frahm // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – С. 4104-4113.
50. Wang Z. Multiscale structural similarity for image quality assessment / Z. Wang, E.P. Simoncelli, A.C. Bovik // The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003 The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. – 2003. – Т. 2. – С. 1398-1402 Vol.2.
51. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric / R. Zhang [и др.] number: arXiv:1801.03924arXiv:1801.03924 [cs]. – arXiv, 2018

## References

1. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. NeRF / B. Mildenhall [и др.] number: arXiv:2003.08934arXiv:2003.08934 [cs]. – arXiv, 2020.
2. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections. NeRF in the Wild / R. Martin-Brualla [и др.] number: arXiv:2008.02268arXiv:2008.02268 [cs]. – arXiv, 2021.
3. Instant Neural Graphics Primitives with a Multiresolution Hash Encoding / T. Müller [и др.] // ACM Transactions on Graphics. – 2022. – Т. 41. – № 4. – pp. 1-15.
4. Levoy M. Light field rendering / M. Levoy, P. Hanrahan // Proceedings of the 23rd annual conference on Computer graphics and interactive techniques : SIGGRAPH '96. – New York, NY, USA: Association for Computing Machinery, 1996. – pp. 31-42.
5. Davis A. Unstructured Light Fields / A. Davis, M. Levoy, F. Durand // Computer Graphics Forum. – 2012. – Т. 31.
6. The Lumigraph / S. Gortler [и др.] // Proc. of SIGGRAPH 96. – 2001. – Т. 96.
7. Waechter M. Let There Be Color! Large-Scale Texturing of 3D Reconstructions / M. Waechter, N. Moehrle, M. Goesele // Computer Vision – ECCV 2014 : Lecture Notes in Computer Science / eds. D. Fleet [et al.]. – Cham: Springer International Publishing, 2014. – pp. 836-850.
8. Unstructured lumigraph rendering / C. Buehler [и др.] // Proceedings of the 28th annual conference on Computer graphics and interactive techniques : SIGGRAPH '01. – New York, NY, USA: Association for Computing Machinery, 2001. – pp. 425-432.
9. Surface light fields for 3D photography / D.N. Wood [и др.] // Proceedings of the 27th annual conference on Computer graphics and interactive techniques : SIGGRAPH '00. – USA: ACM Press/Addison-Wesley Publishing Co., 2000. – pp. 287-296.
10. Learning to Predict 3D Objects with an Interpolation-based Differentiable Renderer / W. Chen [и др.] number: arXiv:1908.01210arXiv:1908.01210 [cs]. – arXiv, 2019.
11. Unsupervised Training for 3D Morphable Model Regression / K. Genova [и др.] number: arXiv:1806.06098arXiv:1806.06098 [cs]. – arXiv, 2018.
12. Soft Rasterizer: A Differentiable Renderer for Image-based 3D Reasoning. Soft Rasterizer / S. Liu [и др.] number: arXiv:1904.01786arXiv:1904.01786 [cs]. – arXiv, 2019.
13. Loper M.M. OpenDR: An Approximate Differentiable Renderer / M.M. Loper, M.J. Black // Computer Vision – ECCV 2014 : Lecture Notes in Computer Science / eds. D. Fleet [et al.]. – Cham: Springer International Publishing, 2014. – OpenDR. – pp. 154-169.
14. Differentiable Monte Carlo ray tracing through edge sampling / T.-M. Li [и др.] // ACM Transactions on Graphics. – 2018. – Т. 37. – № 6. – pp. 222:1-222:11.
15. Kutulakos K.N. A Theory of Shape by Space Carving / K.N. Kutulakos, S.M. Seitz // International Journal of Computer Vision. – 2000. – Vol. 38. – № 3. – pp. 199-218.
16. Seitz S.M. Photorealistic Scene Reconstruction by Voxel Coloring / S.M. Seitz, C.R. Dyer // International Journal of Computer Vision. – 1999. – Vol. 35. – № 2. – pp. 151-173.
17. Szeliski R. Stereo Matching with Transparency and Matting / R. Szeliski, P. Golland // International Journal of Computer Vision. – 1999. – Vol. 32. – № 1. – pp. 45-61.
18. DeepView: View Synthesis with Learned Gradient Descent. DeepView / J. Flynn [и др.] number: arXiv:1906.07316arXiv:1906.07316 [cs, eess]. – arXiv, 2019.

19. Single-image Tomography: 3D Volumes from 2D Cranial X-Rays. Single-image Tomography / P. Henzler [и др.] number: arXiv:1710.04867arXiv:1710.04867 [cs]. – arXiv, 2018.
20. Kar A. Learning a Multi-View Stereo Machine / A. Kar, C. Häne, J. Malik // Advances in Neural Information Processing Systems. – Curran Associates, Inc., 2017. – Т. 30.
21. Local Light Field Fusion: Practical View Synthesis with Prescriptive Sampling Guidelines. Local Light Field Fusion / B. Mildenhall [и др.] number: arXiv:1905.00889arXiv:1905.00889 [cs]. – arXiv, 2019.
22. Penner E. Soft 3D reconstruction for view synthesis / E. Penner, L. Zhang // ACM Transactions on Graphics. – 2017. – Т. 36. – № 6. – pp. 235:1-235:11.
23. Pushing the Boundaries of View Extrapolation with Multiplane Images / P.P. Srinivasan [и др.] number: arXiv:1905.00413arXiv:1905.00413 [cs]. – arXiv, 2019.
24. Multi-view Supervision for Single-view Reconstruction via Differentiable Ray Consistency / S. Tulsiani [и др.] number: arXiv:1704.06254arXiv:1704.06254 [cs]. – arXiv, 2017.
25. Stereo Magnification: Learning View Synthesis using Multiplane Images. Stereo Magnification / T. Zhou [и др.] number: arXiv:1805.09817arXiv:1805.09817 [cs]. – arXiv, 2018.
26. Porter T. Compositing digital images / T. Porter, T. Duff // ACM SIGGRAPH Computer Graphics. – 1984. – Vol. 18. – № 3. – pp. 253-259.
27. DeepVoxels: Learning Persistent 3D Feature Embeddings. DeepVoxels / V. Sitzmann [и др.] number: arXiv:1812.01024arXiv:1812.01024 [cs]. – arXiv, 2019.
28. Neural Volumes: Learning Dynamic Renderable Volumes from Images / S. Lombardi [и др.] // ACM Transactions on Graphics. – 2019. – Т. 38. – Neural Volumes. – № 4. – pp. 1-14.
29. Hartley R. Multiple View Geometry in Computer Vision / R. Hartley, A. Zisserman. – 2. – Cambridge: Cambridge University Press, 2004.
30. Bundle Adjustment — A Modern Synthesis / B. Triggs [et al.] // Vision Algorithms: Theory and Practice : Lecture Notes in Computer Science / eds. B. Triggs, A. Zisserman, R. Szeliski. – Berlin, Heidelberg: Springer, 2000. – pp. 298-372.
31. Snavely N. Photo tourism: exploring photo collections in 3D / N. Snavely, S.M. Seitz, R. Szeliski // ACM Transactions on Graphics. – 2006. – Т. 25. – Photo tourism. – № 3. – pp. 835-846.
32. Reconstructing Rome / S. Agarwal [и др.] // Computer. – 2010. – Т. 43. – № 6. – pp. 40-47.
33. Building Rome on a Cloudless Day / J.-M. Frahm [et al.] // Computer Vision – ECCV 2010 : Lecture Notes in Computer Science / eds. K. Daniilidis, P. Maragos, N. Paragios. – Berlin, Heidelberg: Springer, 2010. – pp. 368-381.
34. Coherent Intrinsic Images from Photo Collections : Proceedings of ACM SIGGRAPH Asia 2012 / P.-Y. Laffont [et al.] // ACM Transactions on Graphics. – 2012. – Vol. 31. – № 6.
35. The Visual Turing Test for Scene Reconstruction / Q. Shan [и др.] // 2013 International Conference on 3D Vision - 3DV 2013 2013 International Conference on 3D Vision - 3DV 2013. – 2013. – pp. 25-32.
36. Learning to Factorize and Relight a City / A. Liu [и др.] // European Conference on Computer Vision (ECCV). – 2020.
37. Augmenting Crowd-Sourced 3D Reconstructions Using Semantic Detections / T. Price [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2018. – pp. 1926-1935.

38. State of the Art on Neural Rendering / A. Tewari [и др.] number: arXiv:2004.03805arXiv:2004.03805 [cs]. – arXiv, 2020.
  39. Image-to-Image Translation with Conditional Adversarial Networks / P. Isola [и др.] number: arXiv:1611.07004arXiv:1611.07004 [cs]. – arXiv, 2018.
  40. LookinGood: Enhancing Performance Capture with Real-time Neural Re-Rendering. LookinGood / R. Martin-Brualla [и др.] number: arXiv:1811.05029arXiv:1811.05029 [cs]. – arXiv, 2018.
  41. Thies J. Deferred Neural Rendering: Image Synthesis using Neural Textures. Deferred Neural Rendering / J. Thies, M. Zollhöfer, M. Nießner number: arXiv:1904.12356arXiv:1904.12356 [cs]. – arXiv, 2019.
  42. Neural Point-Based Graphics / K.-A. Aliev [и др.] number: arXiv:1906.08240arXiv:1906.08240 [cs]. – arXiv, 2020.
  43. DeepStereo: Learning to Predict New Views from the World’s Imagery. DeepStereo / J. Flynn [и др.] number: arXiv:1506.06825arXiv:1506.06825 [cs]. – arXiv, 2015.
  44. Crowdsampling the Plenoptic Function / Z. Li [и др.] number: arXiv:2007.15194arXiv:2007.15194 [cs]. – arXiv, 2020.
  45. Sitzmann V. Scene Representation Networks: Continuous 3D-Structure-Aware Neural Scene Representations. Scene Representation Networks / V. Sitzmann, M. Zollhöfer, G. Wetzstein number: arXiv:1906.01618arXiv:1906.01618 [cs]. – arXiv, 2020.
  46. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains / M. Tancik [и др.] number: arXiv:2006.10739arXiv:2006.10739 [cs]. – arXiv, 2020.
  47. Max N. Optical models for direct volume rendering / N. Max // IEEE Transactions on Visualization and Computer Graphics. – 1995. – Т. 1. – № 2. – pp. 99-108.
  48. Optimizing the Latent Space of Generative Networks / P. Bojanowski [и др.] number: arXiv:1707.05776arXiv:1707.05776 [cs, stat]. – arXiv, 2019.
  49. Schonberger J.L. Structure-From-Motion Revisited / J.L. Schonberger, J.-M. Frahm // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – pp. 4104-4113.
  50. Wang Z. Multiscale structural similarity for image quality assessment / Z. Wang, E.P. Simoncelli, A.C. Bovik // The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003 The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. – 2003. – Т. 2. – pp. 1398-1402 Vol.2.
  51. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric / R. Zhang [и др.] number: arXiv:1801.03924arXiv:1801.03924 [cs]. – arXiv, 2018
-