




COLMAP и BARF

методы конструирования 3D-объектов по фотографиям



Введение в компьютерное зрение и 3D реконструкцию

Компьютерное зрение:

- Цель: Понимание и автоматическая обработка визуальных данных.
- Задачи: Распознавание образов, анализ движения, 3D-реконструкция.

3D-реконструкция:

- Цель: Создание трехмерных моделей из двумерных изображений.
- Почему это важно?



COLMAP и BARF. Введение.

COLMAP

- Программное обеспечение
- Основа: SfM, MVS
- Функции: 3D-реконструкция по набору фотографий, автоматическое обнаружение
- Традиционные методы

BARF

- Метод для улучшения процесса NeRF
- Уникальность: способность работы с неточными данными
- Функции: оптимизация позиций камеры, применение ГО для обработки данных

COLMAP. Подробно.

История и развитие COLMAP

- COLMAP был разработан Яном Шёнбергером (Johannes L. Schönberger) и его командой.
- Начало примерно в 2016 году.
- Цель: создание универсального инструмента для автоматической 3D-реконструкции из изображений





Основные функции и принцип работы.

1. План работы приложения:
 - a. Входные данные: набор изображения объекта (jpeg), входные параметры (необяз.).
 - b. SfM - создание структуры сцены, оценка параметров данных.
 - c. MVS - детальное конструирование 3D модели.
 - d. Выход - 3D-модель объекта (.ply).
2. Принцип обработки:
 - a. Structure-from-Motion (SfM)
 - b. Multi-View Stereo (MVS)

Structure-from-Motion.

SfM позволяет создавать трехмерные структуры из двухмерных изображений.

Обнаружение и сопоставление ключевых точек (SIFT) -> Оценка позиций камеры (RANSAC) -> Триангуляция -> Создание разреженного облака точек -> Итеративная

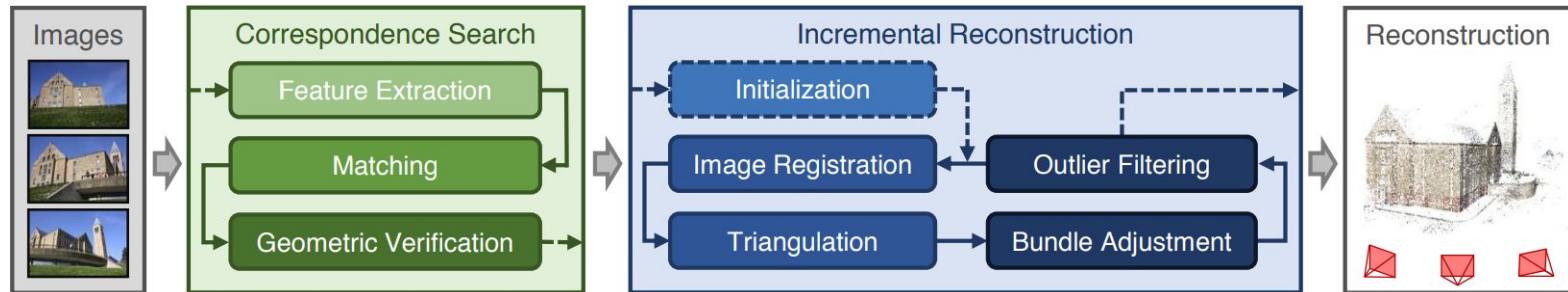


Figure 2. Incremental Structure-from-Motion pipeline.



Связная корректировка (Bundle Adjustment)

ВА - коррекция параметров положения камеры.

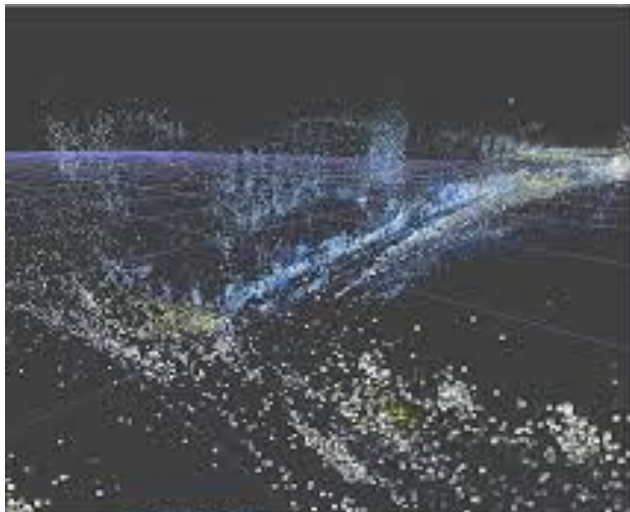
$$E = \sum_j \rho_j \left(\left\| \pi(\mathbf{P}_c, \mathbf{X}_k) - \mathbf{x}_j \right\|_2^2 \right)$$

Функция минимизации

ρ - функция потерь

π - функция проекции

Примеры работы MfS



100_7100.JPG



100_7101.JPG



100_7102.JPG



100_7103.JPG



100_7104.JPG



100_7105.JPG



100_7106.JPG



100_7107.JPG



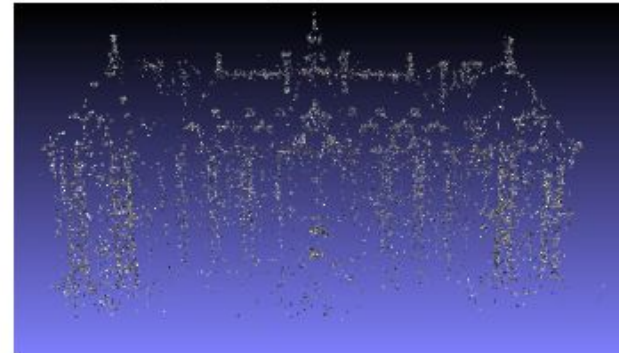
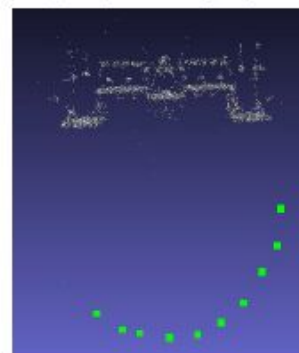
100_7108.JPG



100_7109.JPG



100_7110.JPG





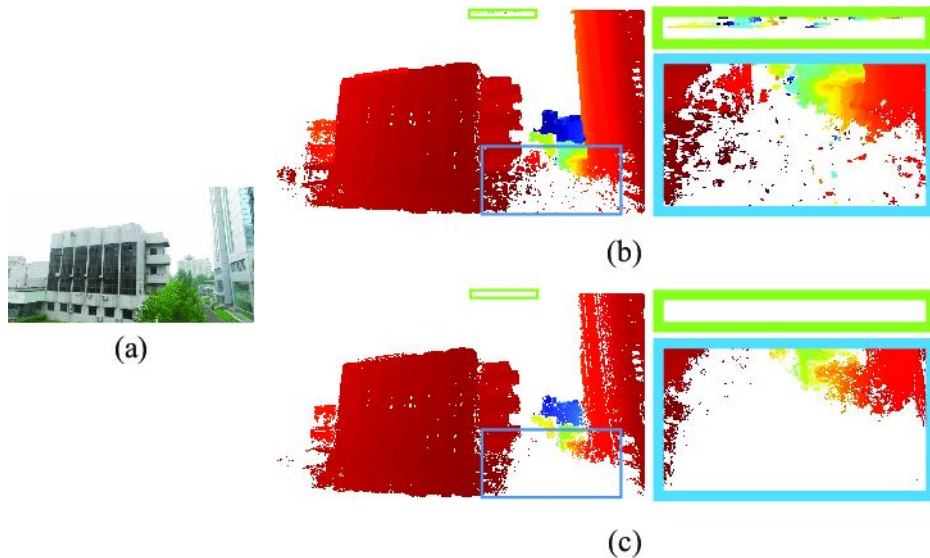
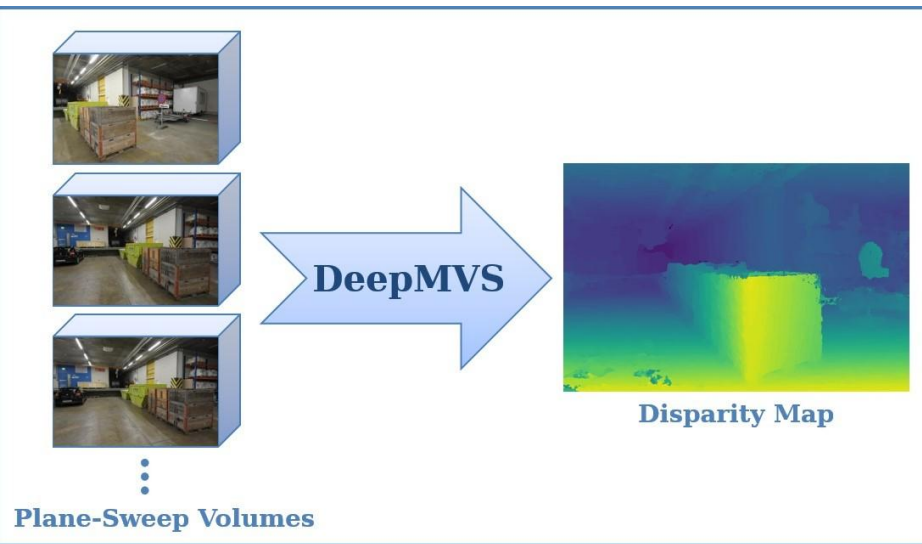
Multi-View Stereo

MVS, или **мульти-вью стереоскопия**, это метод, используемый для создания детализированных трехмерных моделей из множества двумерных изображений. (Следует за SfM)

Принцип работы:

Входные данные (изображения + данные о камере) -> оценка глубины изображений, составления глубинных карт -> Триангуляция по разреженному облаку (MfS) и картам -> Создание плотного облака -> устранение шума и ошибок.

Примеры работы MVS.



Результаты работы COLMAP



BARF. Подробно.



NeRF. “Нейронный” подход к задаче.

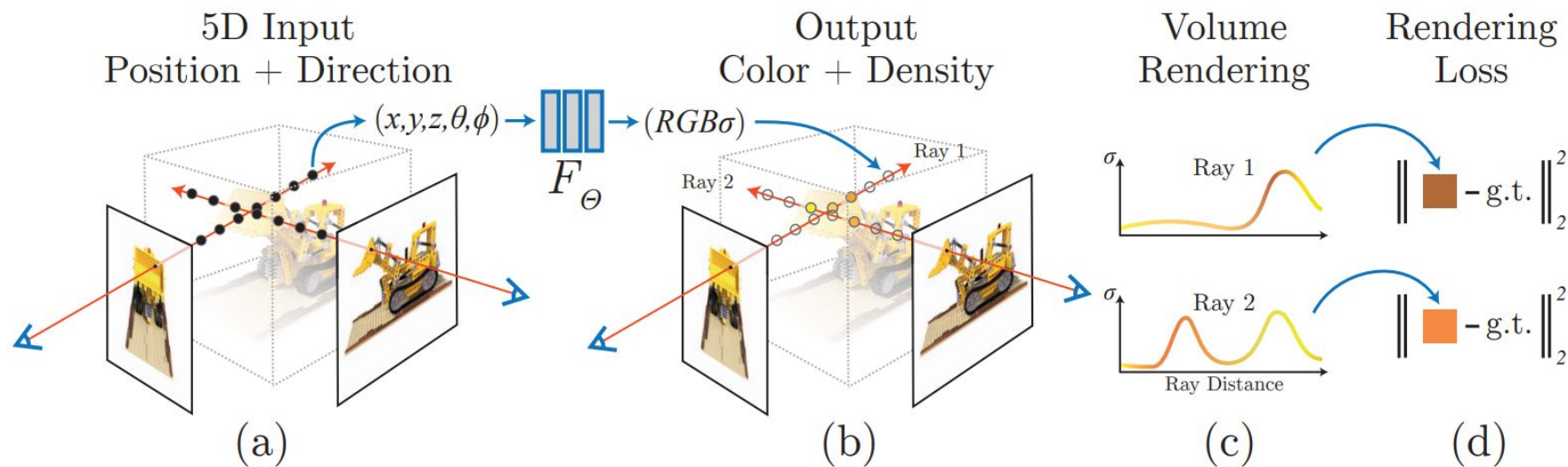
Neural Radiance Fields (NeRF) — это метод, который использует глубокое обучение для синтеза фотореалистичных трехмерных сцен из набора двумерных изображений.

Процесс работы NeRF:

1. Вход. данные: изображения, подробная характеристика камер (или SfM)
2. Проецирование лучей, выбор точек на этих лучах
3. Обучение сети.
4. Объемный рендеринг (Volume Rendering)
5. 3D модель.

Принцип работы модели.

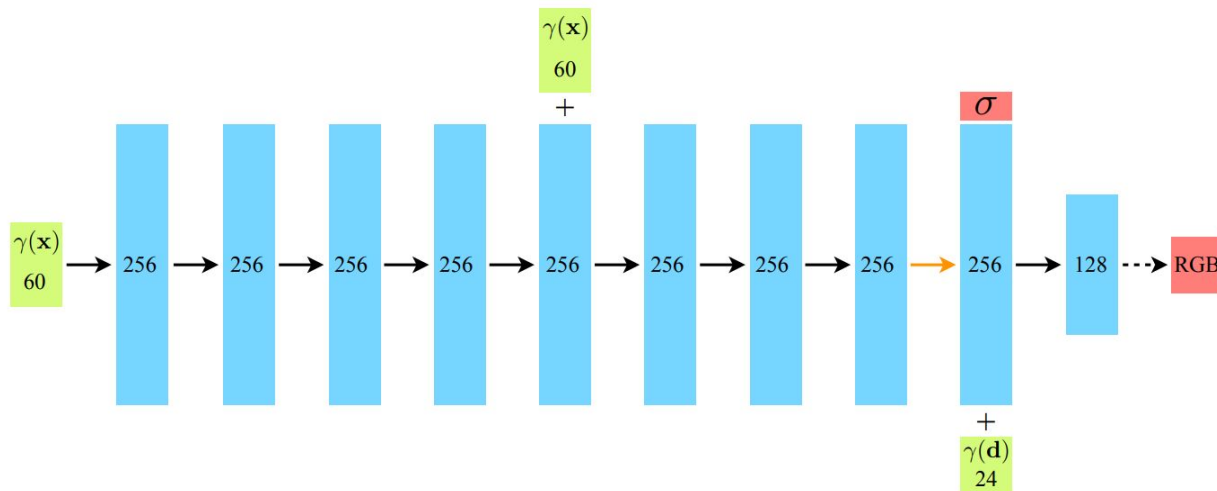
$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left[\left\| \hat{C}_c(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 + \left\| \hat{C}_f(\mathbf{r}) - C(\mathbf{r}) \right\|_2^2 \right]$$



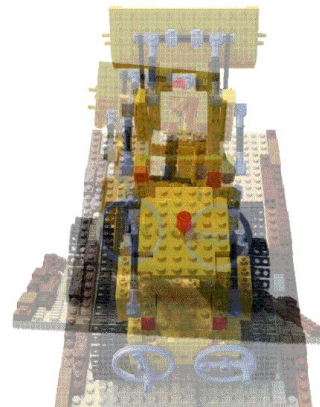
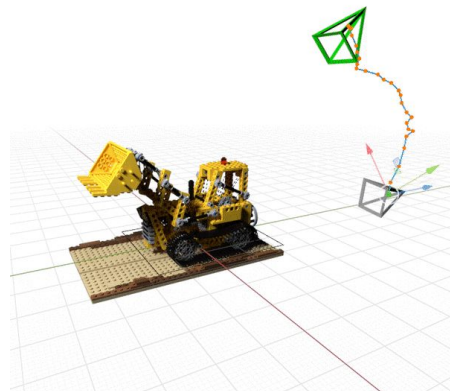
Нейросеть в NeRF.

В модели используется **полносвязные нейронные сети** (не сверточные).

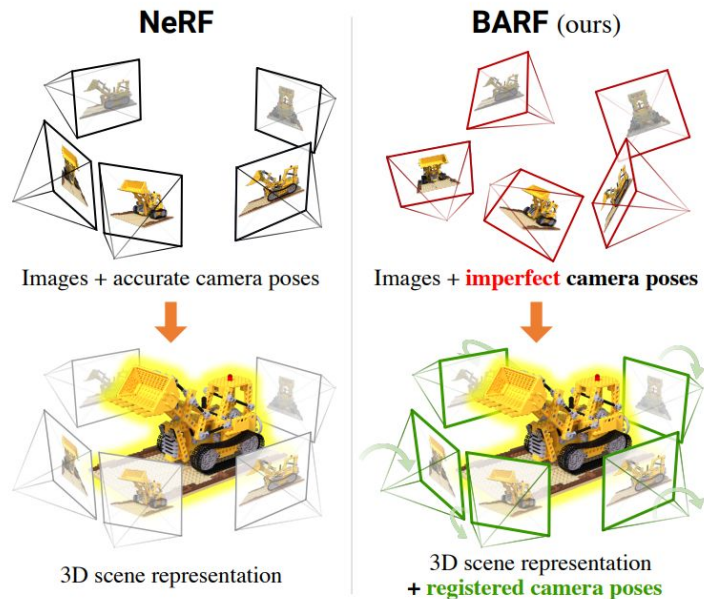
Вход - (x, y, z, θ, ϕ) , выход - (RGB, σ)



Примеры



BARF, как новый уровень для NeRF



Bundle-Adjusting Neural Radiance Fields (BARF) является усовершенствованием оригинальной модели NeRF.

Почему NeRF не идеален? Основная проблема - необходимость точных входных данных.

Отличия. Чего хотим добиться?

Было:

$$\min_{\Theta} \sum_{i=1}^{\text{frames } M} \sum_{\mathbf{u}} \left\| \hat{\mathcal{I}}(\mathbf{u}; \mathbf{p}_i, \Theta) - \mathcal{I}_i(\mathbf{u}) \right\|_2^2$$

Legend: \mathbf{p}_i (camera pose), Θ (network params.), $\mathcal{I}_i(\mathbf{u})$ (RGB)

Хотим:

$$\min_{\mathbf{p}_1, \dots, \mathbf{p}_M, \Theta} \sum_{i=1}^{\text{frames } M} \sum_{\mathbf{u}} \left\| \hat{\mathcal{I}}(\mathbf{u}; \mathbf{p}_i, \Theta) - \mathcal{I}_i(\mathbf{u}) \right\|_2^2$$

Legend: $\mathbf{p}_1, \dots, \mathbf{p}_M$ (camera pose), Θ (network params.), $\mathcal{I}_i(\mathbf{u})$ (RGB)



Решение проблемы.

Bundle-Adjusting.

$$\min_{\mathbf{p}_1, \dots, \mathbf{p}_M, \Theta} \sum_{i=1}^M \sum_{\mathbf{u}} \|\hat{\mathcal{I}}(\mathbf{u}; \mathbf{p}_i, \Theta) - \mathcal{I}_i(\mathbf{u})\|_2^2$$

$\{\mathbf{p}_i\}_{i=1}^M$ -параметры камеры

$\{\mathcal{I}_i\}_{i=1}^M$ -исх. фотографии

Так же как и в SfM тут присутствует идея подбора параметров положения камеры во время “обучения”.

Решение проблемы.

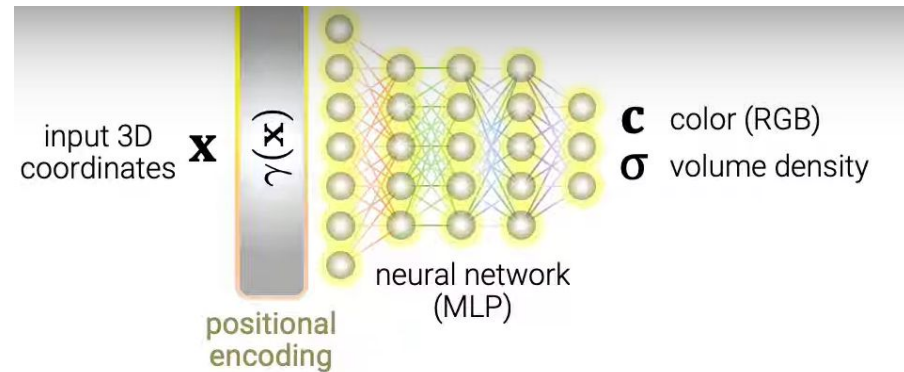
Изменение positional encoding:

$$\gamma_k(\mathbf{x}) = w_k \cdot [\cos(2^k \pi \mathbf{x}), \sin(2^k \pi \mathbf{x})]$$

$\in [0,1]$

$$\text{where } w_k(\alpha) = \begin{cases} 0 & \text{if } \alpha < k \\ \frac{1 - \cos((\alpha - k)\pi)}{2} & \text{if } 0 \leq \alpha - k < 1 \\ 1 & \text{if } \alpha - k \geq 1 \end{cases}$$

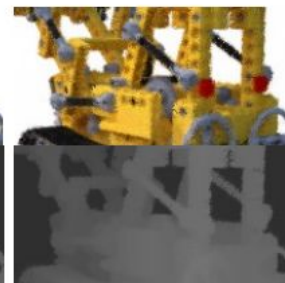
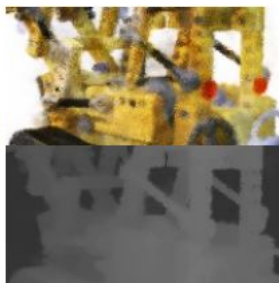
$\alpha \in [0, L]$ is a controllable parameter



Пример.



lego



ficus



ground truth

full pos. enc.

w/o pos. enc.

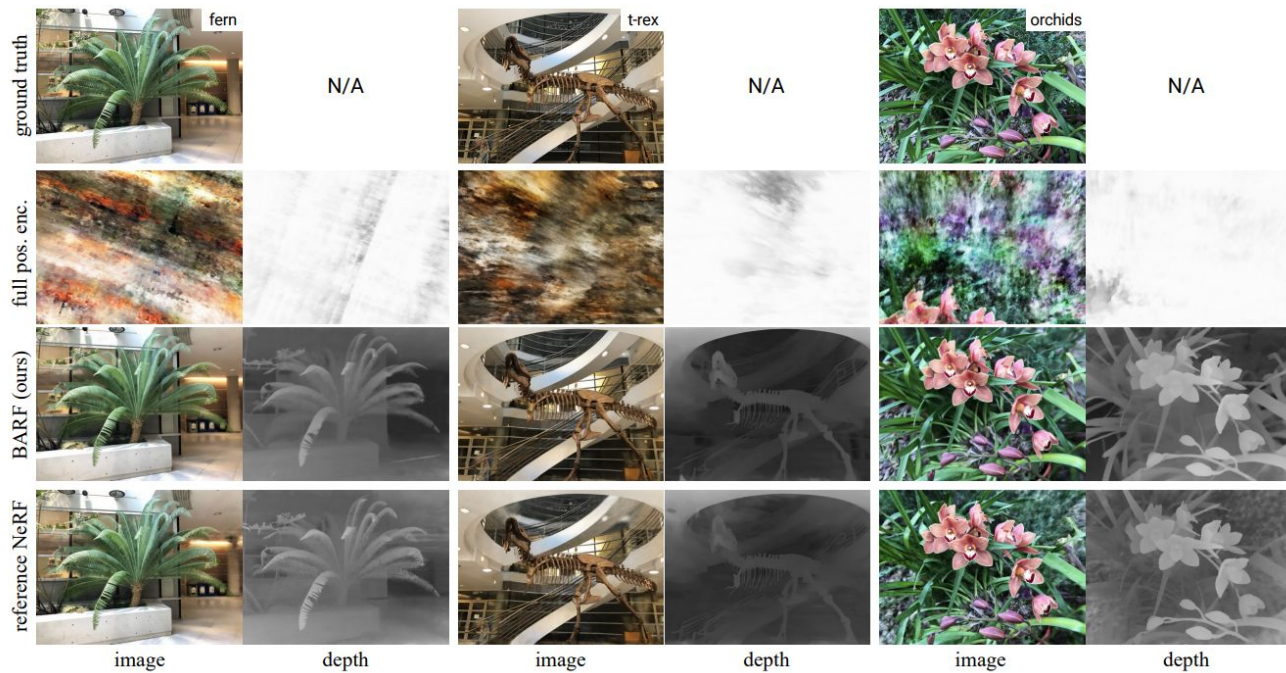
BARF (ours)

reference NeRF

Результаты.

Scene	Camera pose registration						View synthesis quality											
	Rotation (°) ↓			Translation ↓			PSNR ↑				SSIM ↑				LPIPS ↓			
	full pos.enc.	w/o pos.enc.	BARF	full pos.enc.	w/o pos.enc.	BARF	full pos.enc.	w/o pos.enc.	BARF	ref. NeRF	full pos.enc.	w/o pos.enc.	BARF	ref. NeRF	full pos.enc.	w/o pos.enc.	BARF	ref. NeRF
Chair	7.186	0.110	0.096	16.638	0.555	0.428	19.02	30.22	31.16	31.91	0.804	0.942	0.954	0.961	0.223	0.065	0.044	0.036
Drums	3.208	0.057	0.043	7.322	0.255	0.225	20.83	23.56	23.91	23.96	0.840	0.893	0.900	0.902	0.166	0.116	0.099	0.095
Ficus	9.368	0.095	0.085	10.135	0.430	0.474	19.75	25.58	26.26	26.68	0.836	0.922	0.934	0.941	0.182	0.070	0.058	0.051
Hotdog	3.290	0.225	0.248	6.344	1.122	1.308	28.15	34.00	34.54	34.91	0.923	0.967	0.970	0.973	0.083	0.040	0.032	0.029
Lego	3.252	0.108	0.082	4.841	0.391	0.291	24.23	26.35	28.33	29.28	0.876	0.880	0.927	0.942	0.102	0.112	0.050	0.037
Materials	6.971	0.845	0.844	15.188	2.678	2.692	16.51	26.86	27.84	28.48	0.747	0.926	0.936	0.944	0.294	0.068	0.058	0.049
Mic	10.554	0.081	0.071	22.724	0.356	0.301	15.10	30.93	31.18	31.98	0.788	0.968	0.969	0.971	0.334	0.050	0.048	0.044
Ship	5.506	0.095	0.075	7.232	0.354	0.326	22.12	26.78	27.50	28.00	0.755	0.833	0.849	0.858	0.255	0.175	0.132	0.118
Mean	6.167	0.202	0.193	11.303	0.768	0.756	22.12	26.78	27.50	29.40	0.821	0.917	0.930	0.936	0.205	0.087	0.065	0.057

Результаты.





Итоги. Сравнение.

- BARF - нейронные сети для воссоздания трехмерных изображений. COLMAP - традиционные алгоритмы компьютерного зрения для 3D-реконструкции.
- BARF фокусируется на улучшении качества визуализации в NeRF
COLMAP - на создание точных 3D-моделей сцен.
- BARF способен работать с менее точными данными о камере,
COLMAP требует более точной информации для эффективной реконструкции.
- BARF чаще применяется в исследовательских и академических проектах.
COLMAP применяется в более широком диапазоне практических приложений.

.



Материалы

- Официальный репозиторий COLMAP - <https://colmap.github.io/>
- Статья о MfS - <https://demuc.de/papers/schoenberger2016sfm.pdf>
- Статья о MVS - <https://demuc.de/papers/schoenberger2016mvs.pdf>
- Пример работы COLMAP - https://www.youtube.com/watch?v=Zm1mkOi9_1c
- Статья о NeRF - <https://arxiv.org/pdf/2003.08934.pdf>
- Статья о BARF-
[https://openaccess.thecvf.com/content/ICCV2021/papers/Lin BARF Bundle-Adjusting Neural Radiance Fields ICCV 2021 paper.pdf](https://openaccess.thecvf.com/content/ICCV2021/papers/Lin_BARF_Bundle-Adjusting_Neural_Radiance_Fields_ICCV_2021_paper.pdf)