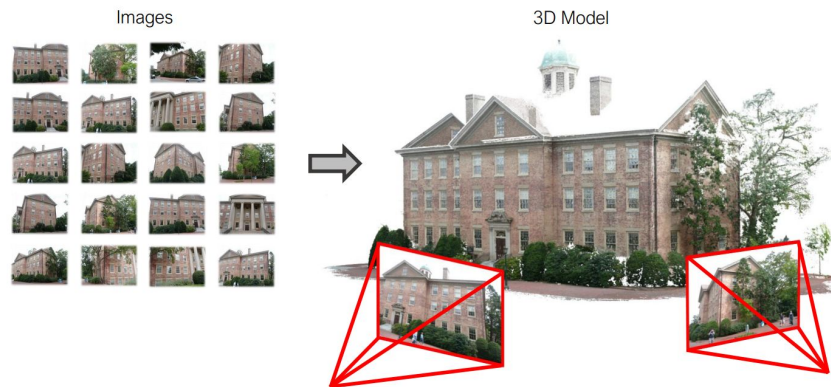


COLMAP + BARF



COLMAP

- Система для реконструирования
- Structure from Motion (SfM) или Multi-view Stereo (MvS)
- SOTA в 2016 году



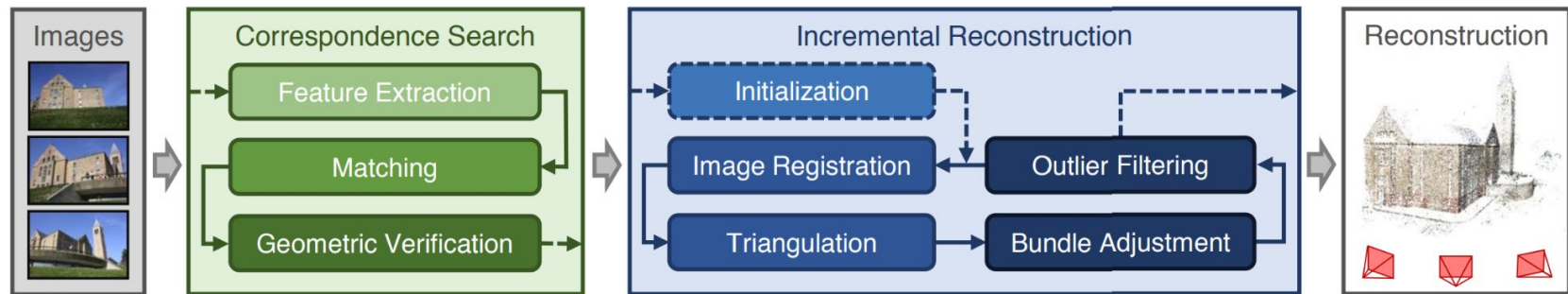
Structure from Motion

- Проблемы



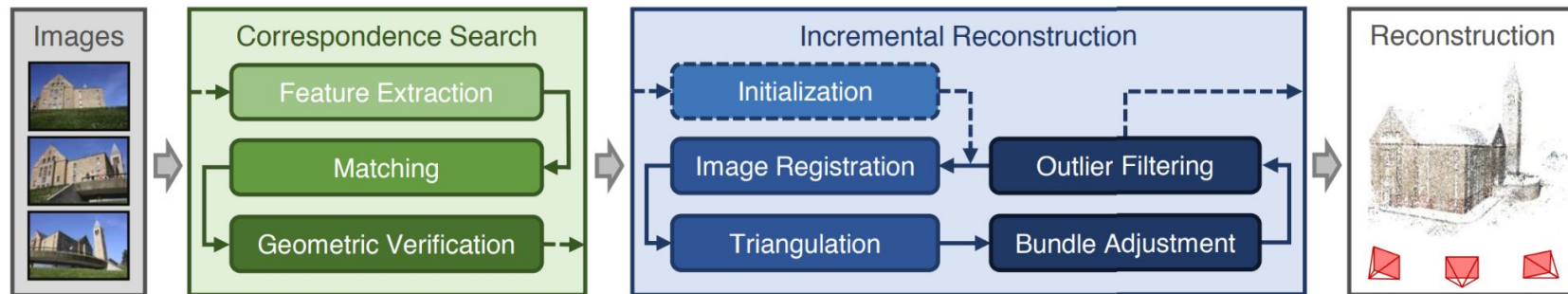
Structure from Motion

- Используется Incremental SfM
- Аккуратная инициализация двумя изображениями
- Далее последовательно:
 - регистрация нового изображения
 - триангуляция точек сцены
 - отсечение выбросов
 - улучшение реконструкции с помощью **Bundle Adjustment (BA)**



Поиск совпадений

- Вычисление признаков
- Сравнение
- Геометрическая верификация



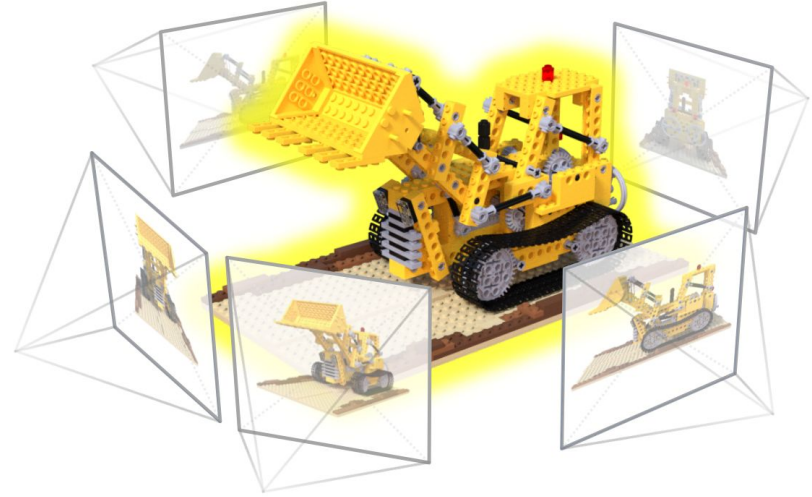
Вычисление признаков

- Множество (позиция, дескриптор)
- Признаки должны быть устойчивы к расположению



Сравнение (матчинг)

- Наивное решение - сравнить каждый признак с каждым
- $O(N^2 M^2)$
- Получаем множество упорядоченных пар совпавших изображений, а также их общие признаки



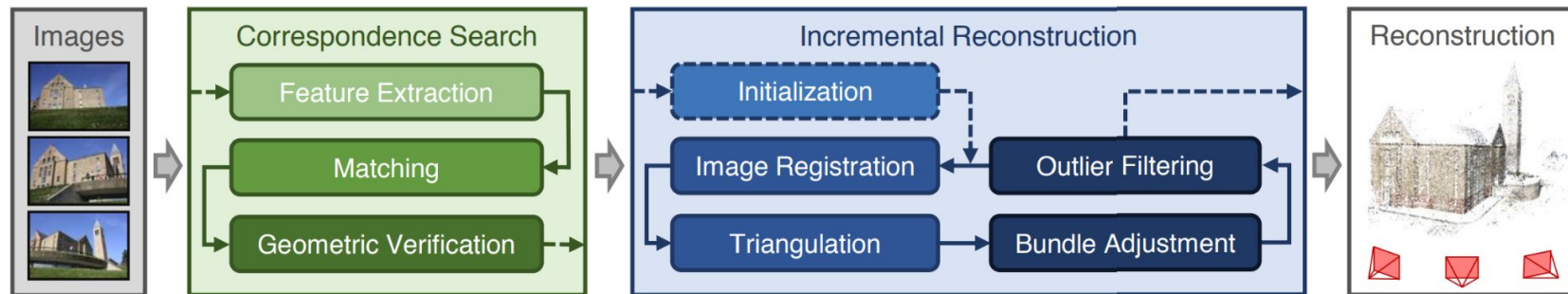
Сравнение (матчинг)

- Строим перспективное преобразование, переводящее одну картинку в другую
- В результате получаем отфильтрованные множества с предыдущего этапа



Последовательное реконструирование

- Инициализация
- Добавление изображения
- Триангуляция
- Bundle Adjustment



Инициализация

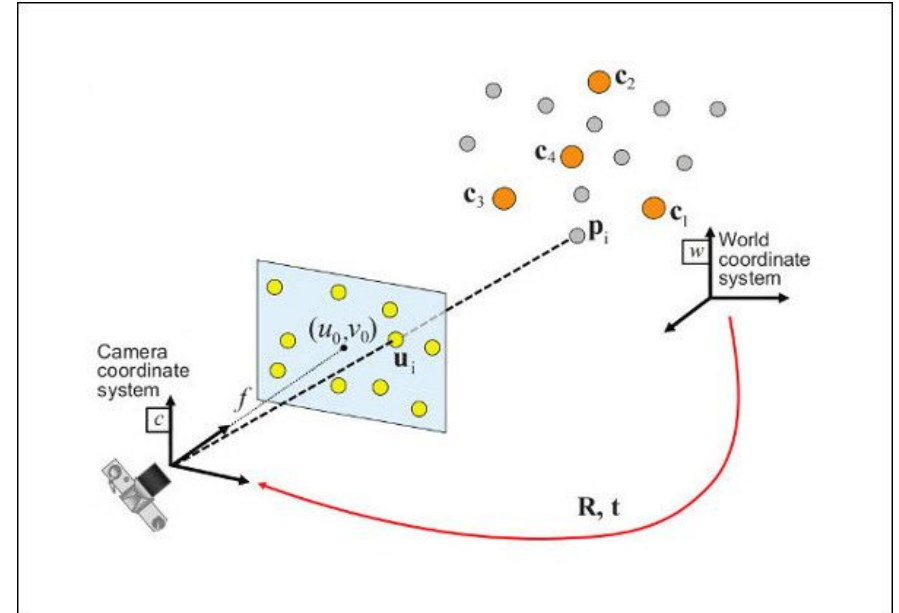
- Выбираем две стартовые точки
- Важно выбрать хорошие точки, иначе реконструкция может не восстановится от плохой инициализации

Добавление изображения

- Лучше всего подходят локации, которым соответствует более связная часть графа и где было больше камер

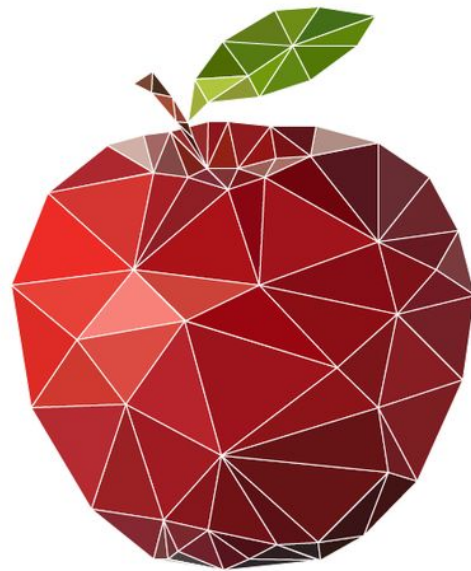
Добавление изображения

- Чтобы понять положение нового изображения, нужно решить Perspective-n-Point задачу
- Получаем позицию камеры и для некалиброванных камер их внутренние параметры



Триангуляция

- Новое изображение проверяет уже наблюдаемые точки триангуляции и может добавить в триангуляцию новые



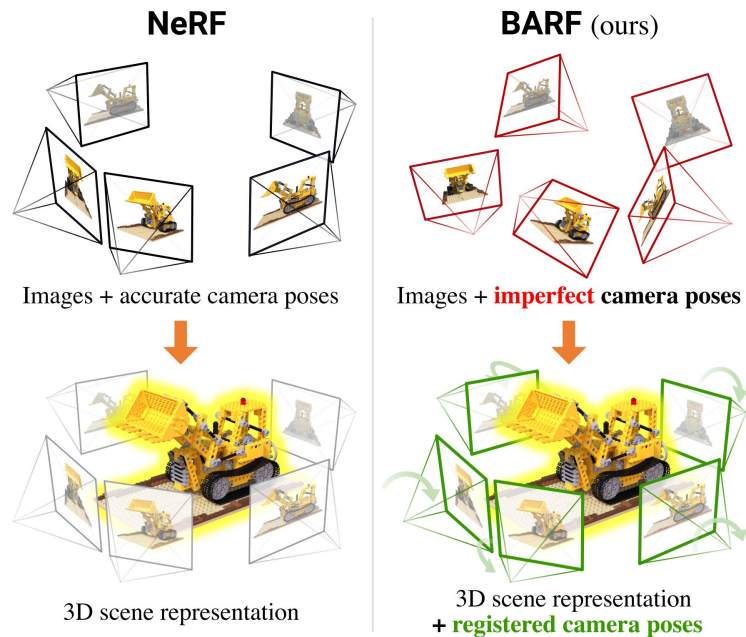
Bundle Adjustment

- Оптимизируем совместную функцию ошибки
- Без этого улучшения модель быстро скатывается в невосстановимое состояние

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

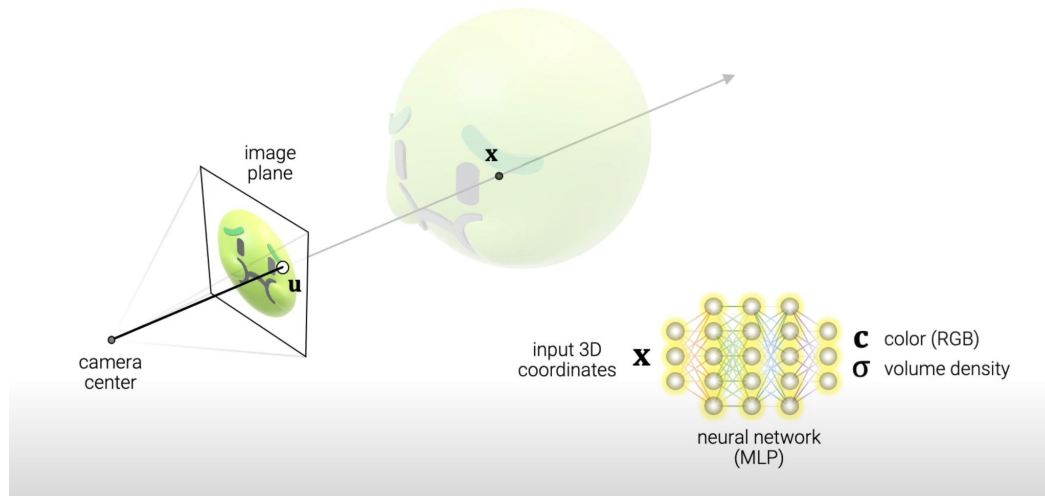
BARF 🤨

- Bundle-Adjusting Neural Radiance Fields
- Улучшение NeRF сети с помощью



Про NeRF

- Предсказываем цвет точки и ее прозрачность для каждой точки на луче
- Далее, совмещая точки на луче, получаем цвет пикселя
- Получили алгоритм объемного рендеринга



Про NeRF

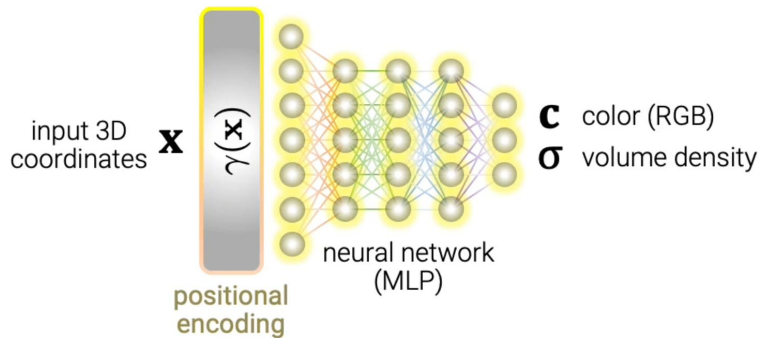
- В качестве функции потерь используется обычная L2 метрика между полученной и наблюдаемой цветностью пикселей
- Возникает проблема, когда позиции камер найдены неаккуратно или вообще неизвестны

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

RGB (rendered) camera pose network params. RGB

Улучшаем NeRF

- Будем использовать дополнительную функцию для преобразования позиции
- В качестве такой функции можно взять преобразование в синусоиды с экспоненциально уменьшающейся длиной волны



$$\gamma(\mathbf{x}) = \begin{bmatrix} \gamma_0(\mathbf{x}) \\ \gamma_1(\mathbf{x}) \\ \gamma_2(\mathbf{x}) \\ \vdots \\ \gamma_{L-1}(\mathbf{x}) \end{bmatrix}$$

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

Улучшаем NeRF

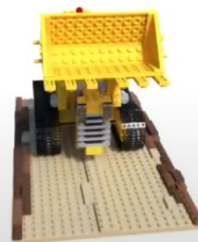
- Получили улучшение детализации мелких объектов



ground truth



NeRF **w/o** pos. enc.



NeRF **w/** pos. enc.

Улучшаем NeRF

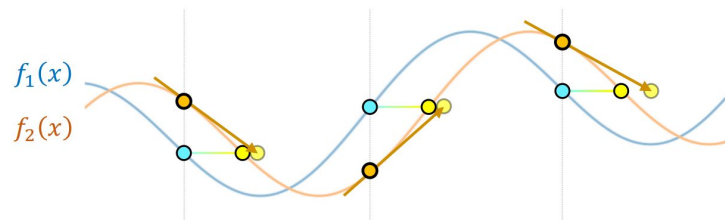
- С другой стороны производная такой функции будет значительно перевешивать в сторону более частых волн

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

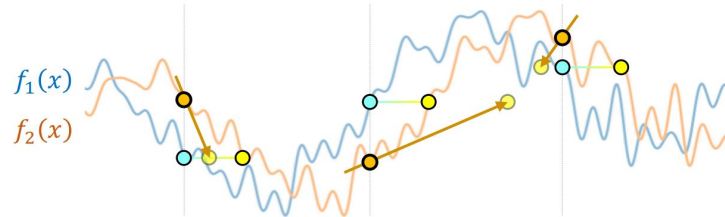


Улучшаем NeRF

- Гладкие сигналы более предпочтительны для геометрического выравнивания



(a) smooth signal



(b) complex signal

BARF

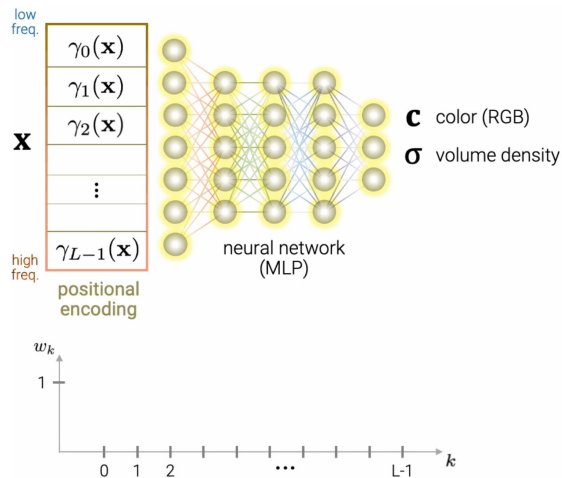
- Добавляем специальный коэффициент для активации волн

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

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

BARF

- Добавляем специальный коэффициент для активации волн

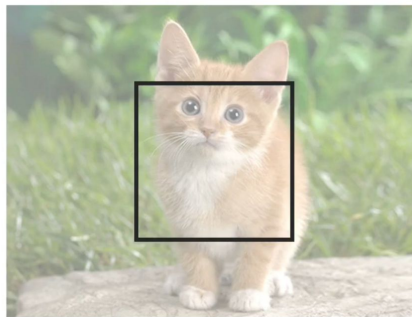


$$\gamma_k(\mathbf{x}; \alpha) = w_k(\alpha) \cdot [\cos(2^k \pi \mathbf{x}), \sin(2^k \pi \mathbf{x})]$$

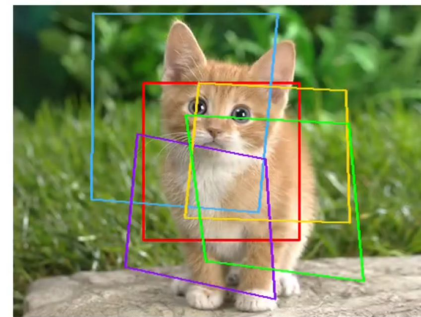
$$w_k(\alpha) = \begin{cases} 0 & \text{if } \alpha < k \\ \frac{1 - \cos((\alpha - k)\pi)}{2} & \text{if } 0 \leq \alpha - k < 1 \\ 1 & \text{if } \alpha - k \geq 1 \end{cases}$$

BARF

- Смотрим, как работает BARF



Initialize all patches to the center

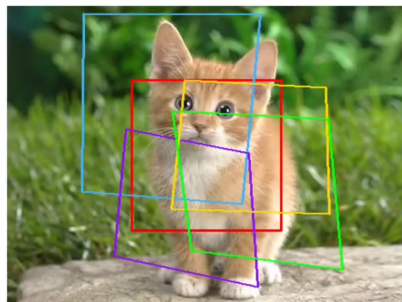


(ground-truth warps)

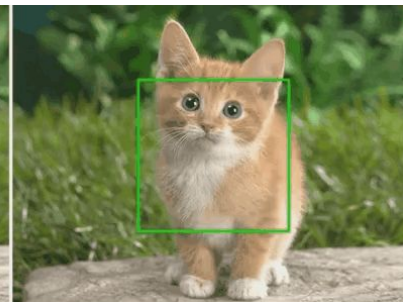
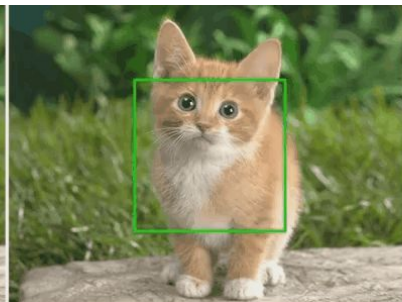
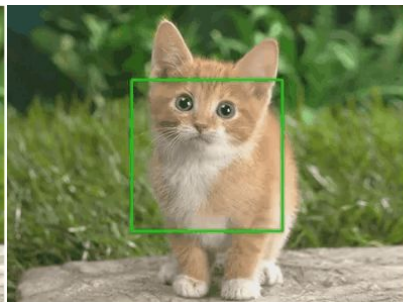


BARF

- Смотрим, как работает BARF



(ground-truth warps)



BARF

- Смотрим, как работает BARF



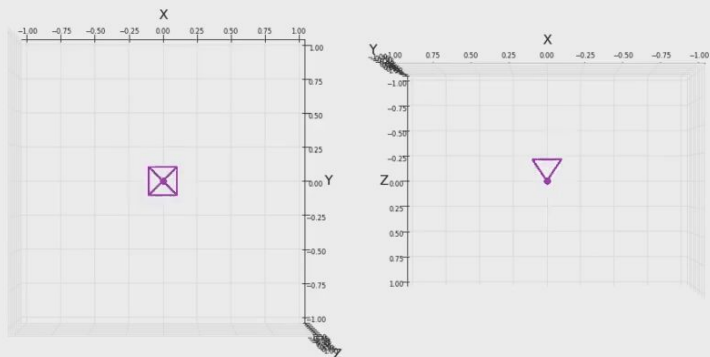
example sequence: orchids (23 frames)



forward-facing view

epoch 0

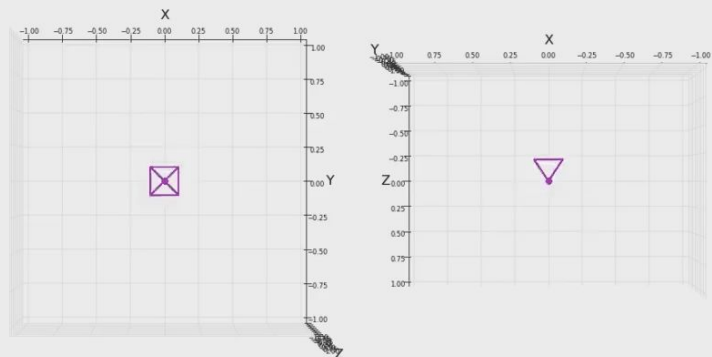
top-down view



forward-facing view

epoch 0

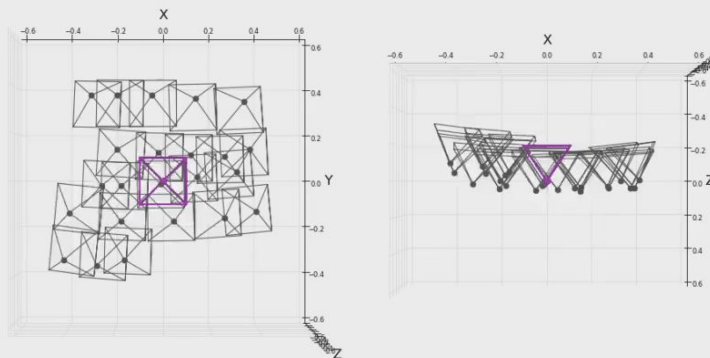
top-down view



forward-facing view

epoch 0

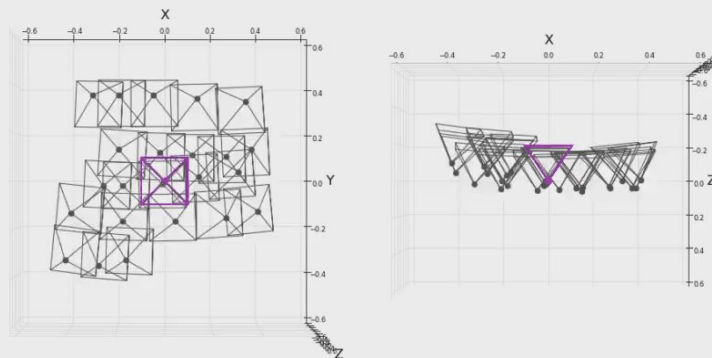
top-down view



forward-facing view

epoch 0

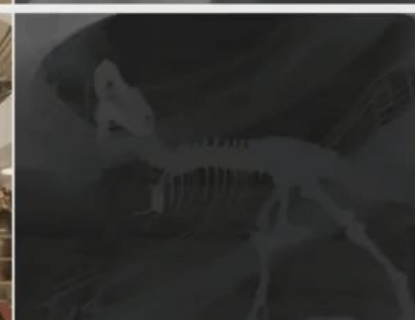
top-down view



Обычное преобразование позиции



BARF



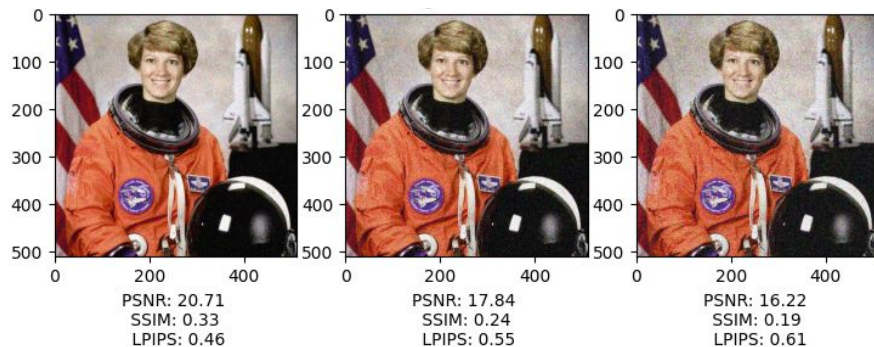
Сравнение тестов BARF

- Сравнение ошибок углов и сдвигов обычного метода с BARF, где за истинное значение приняты позиции из COLMAP

Scene	Camera pose registration				
	Rotation (°) ↓		Translation ↓		
	full pos.enc.	BARF	full pos.enc.	BARF	
Fern	74.452	0.191	30.167	0.192	
Flower	2.525	0.251	2.635	0.224	
Fortress	75.094	0.479	33.231	0.364	
Horns	58.764	0.304	32.664	0.222	
Leaves	88.091	1.272	13.540	0.249	
Orchids	37.104	0.627	20.312	0.404	
Room	173.811	0.320	66.922	0.270	
T-rex	166.231	1.138	53.309	0.720	
Mean	84.509	0.573	31.598	0.331	

Сравнение тестов BARF

- Сравнение метрик изображений обычного метода с BARF
- Добавим обычный NeRF с позициями из COLMAP



View synthesis quality								
PSNR ↑			SSIM ↑			LPIPS ↓		
full pos. enc.	BARF	ref. NeRF	full pos. enc.	BARF	ref. NeRF	full pos. enc.	BARF	ref. NeRF
9.81	23.79	23.72	0.187	0.710	0.733	0.853	0.311	0.262
17.08	23.37	23.24	0.344	0.698	0.668	0.490	0.211	0.244
12.15	29.08	25.97	0.270	0.823	0.786	0.807	0.132	0.185
8.89	22.78	20.35	0.158	0.727	0.624	0.805	0.298	0.421
9.64	18.78	15.33	0.067	0.537	0.306	0.782	0.353	0.526
9.42	19.45	17.34	0.085	0.574	0.518	0.806	0.291	0.307
10.78	31.95	32.42	0.278	0.940	0.948	0.871	0.099	0.080
10.48	22.55	22.12	0.158	0.767	0.739	0.885	0.206	0.244
11.03	23.97	22.56	0.193	0.722	0.665	0.787	0.238	0.283
COLMAP poses			COLMAP poses			COLMAP poses		

Сравнение тестов BARF

- BARF легко использовать
- Достаточно взять фотографии с телефона



COLMAP **fails!** (untextured)

В итоге

- COLMAP – Structure from Motion геометрическая система для распознавания позиций камеры и моделирования 3D изображения по большому набору картинок ([статья и материалы](#))
- BARF – улучшение для NeRF в случае, когда позиции камер неизвестны. Достигается с помощью оптимизации “от грубого к точному” ([статья](#) и [видео](#))