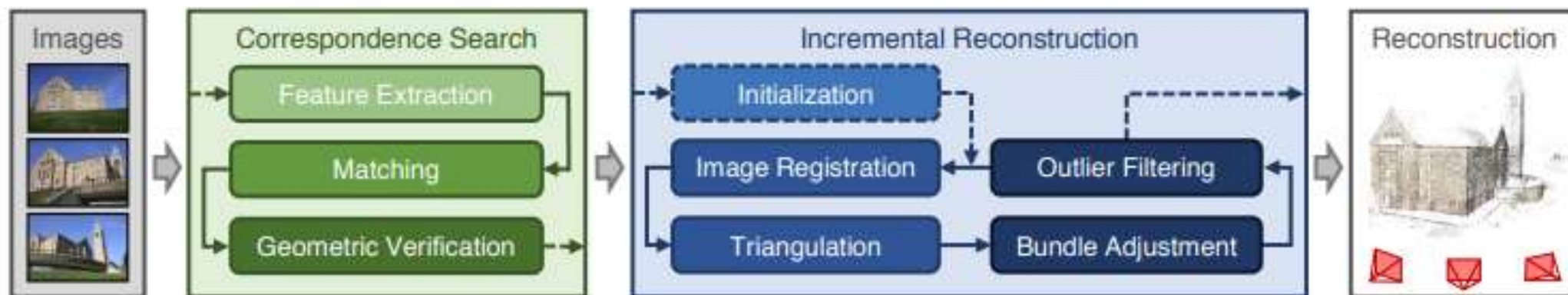


COLMAP + BARF

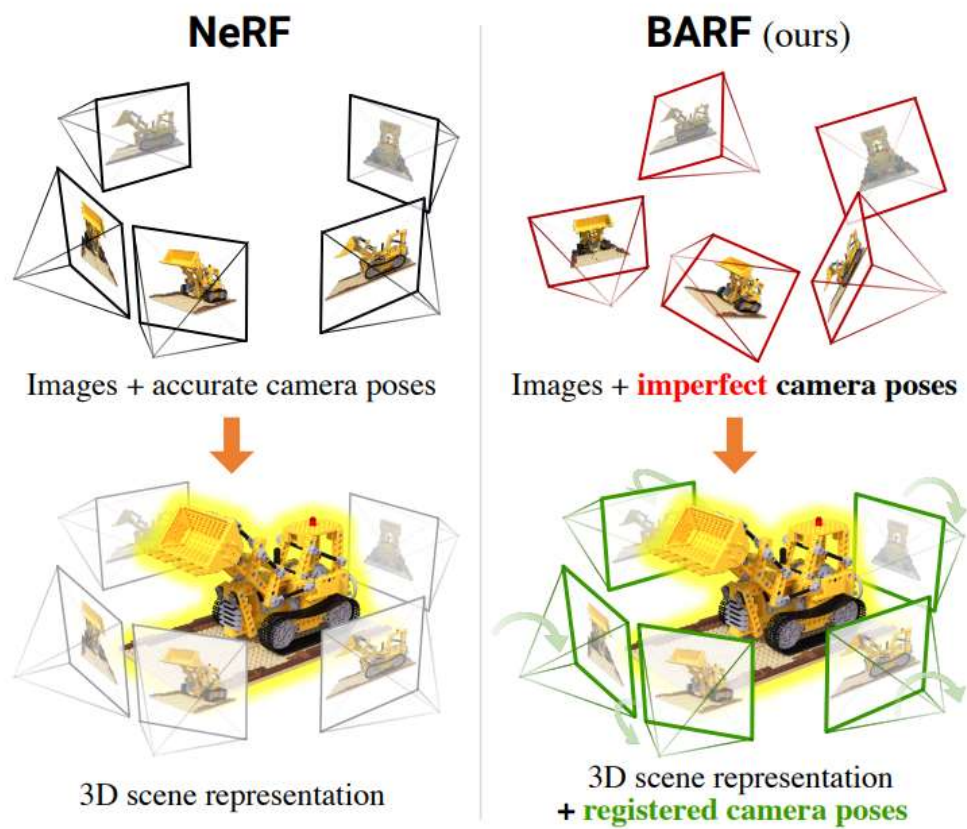
Бекоев Максим

COLMAP

Structure-from-Motion (SfM) – геометрический алгоритм для реконструкции 3D – сцены по набору её изображений



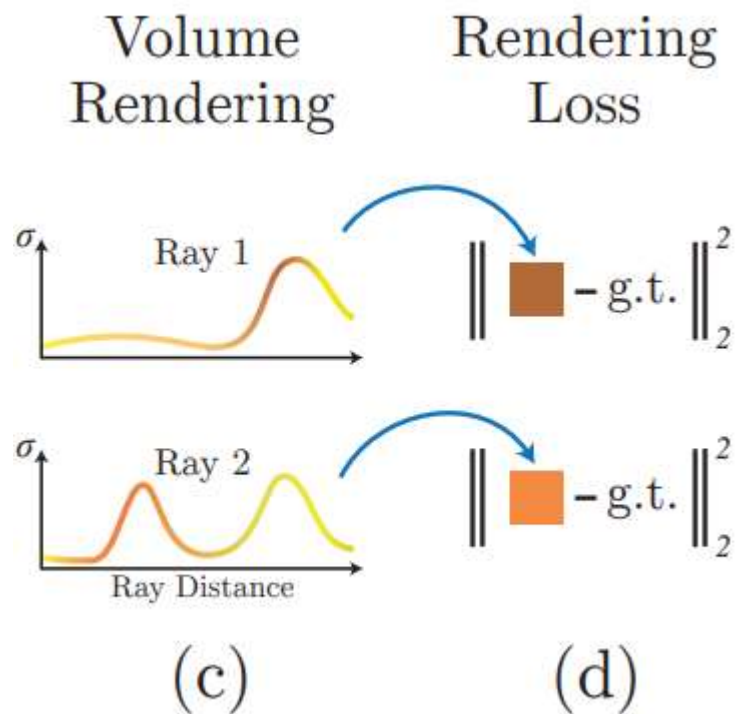
Главный недостаток NeRF – нам нужны точные координаты
расположения камер



Давайте их координаты тоже бросим в
нейронку и будем оптимизировать
вместе с другими весами

Получаем BARF

Хотим считать функцию потерь



g.t. – настоящее значение пикселя

Давайте для каждой картинке построим проекцию предсказанной 3D сцены на соответствующую камеру и посчитаем попиксельно разность

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

Как получить проекцию предсказанной сцены на исходный снимок?

$$\hat{\mathcal{I}}(\mathbf{u}) = \int_{z_{\text{near}}}^{z_{\text{far}}} T(\mathbf{u}, z) \sigma(z\bar{\mathbf{u}}) \mathbf{c}(z\bar{\mathbf{u}}) dz$$

$$T(\mathbf{u}, z) = \exp\left(-\int_{z_{\text{near}}}^z \sigma(z'\bar{\mathbf{u}}) dz'\right)$$

$$\hat{\mathcal{I}}(\mathbf{u}; \mathbf{p}) = g\left(f(\mathcal{W}(z_1\bar{\mathbf{u}}; \mathbf{p}); \Theta), \dots, f(\mathcal{W}(z_N\bar{\mathbf{u}}; \mathbf{p}); \Theta)\right)$$

Positional Encoding

$$\gamma(\mathbf{x}) = [\mathbf{x}, \gamma_0(\mathbf{x}), \gamma_1(\mathbf{x}), \dots, \gamma_{L-1}(\mathbf{x})] \in \mathbb{R}^{3+6L}$$

$$\gamma_k(\mathbf{x}) = [\cos(2^k \pi \mathbf{x}), \sin(2^k \pi \mathbf{x})] \in \mathbb{R}^6$$

Какая возникает проблема?

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

$$y = f(g_1(x), g_2(x), \dots, g_n(x)) \Rightarrow$$
$$\frac{dy}{dx} = \sum_{i=1}^n \frac{\partial f}{\partial g_i} \cdot \frac{dg_i}{dx}$$

Градиенты по разным гамма некогерентны, то есть направлены кто в лес, кто по дрова, и могут гасить друг друга!

Решение – добавим в гамму весовой коэффициент

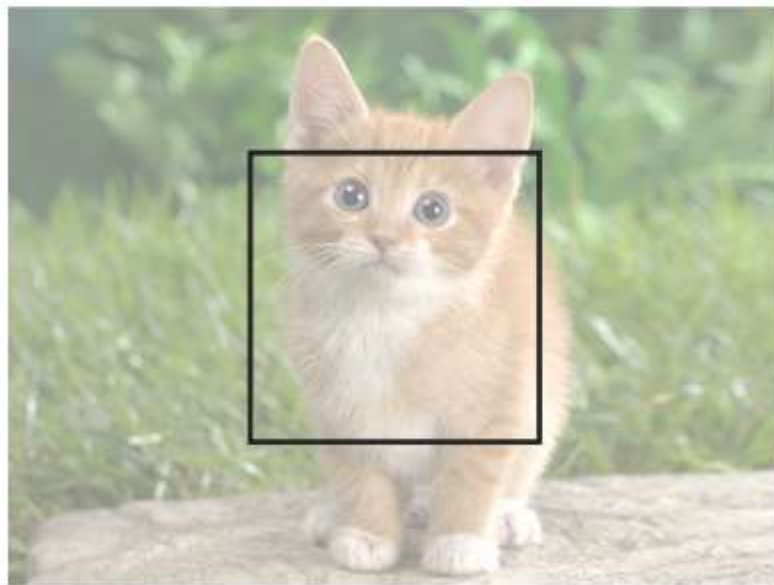
$$\gamma_k(\mathbf{x}; \alpha) = w_k(\alpha) \cdot [\cos(2^k \pi \mathbf{x}), \sin(2^k \pi \mathbf{x})]$$

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

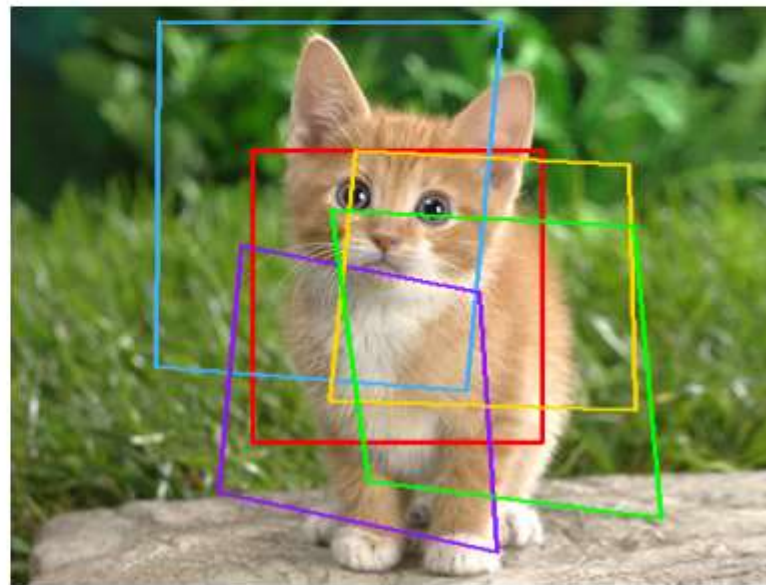
Результаты для двумерного аналога задачи



(a) image patches given for optimization

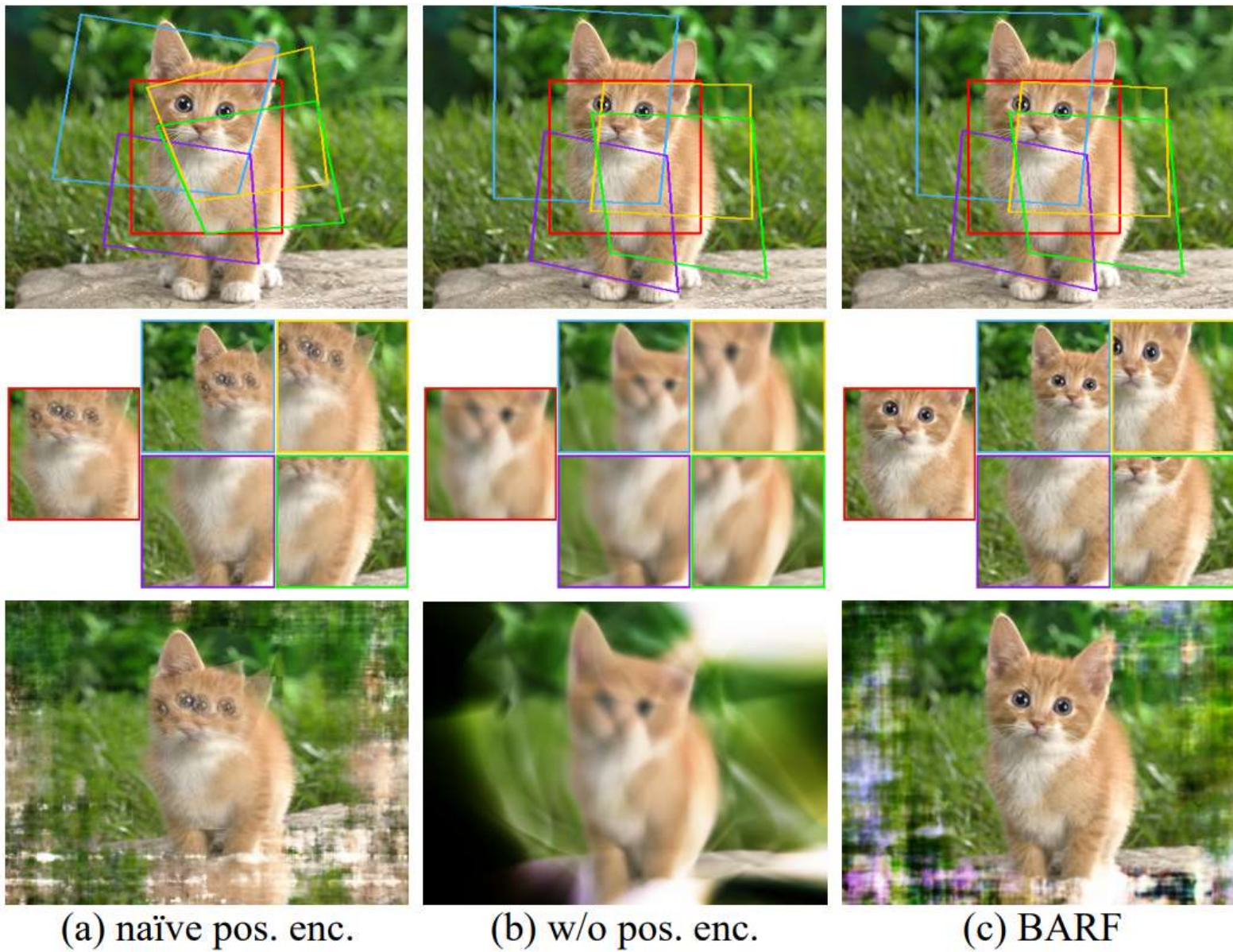


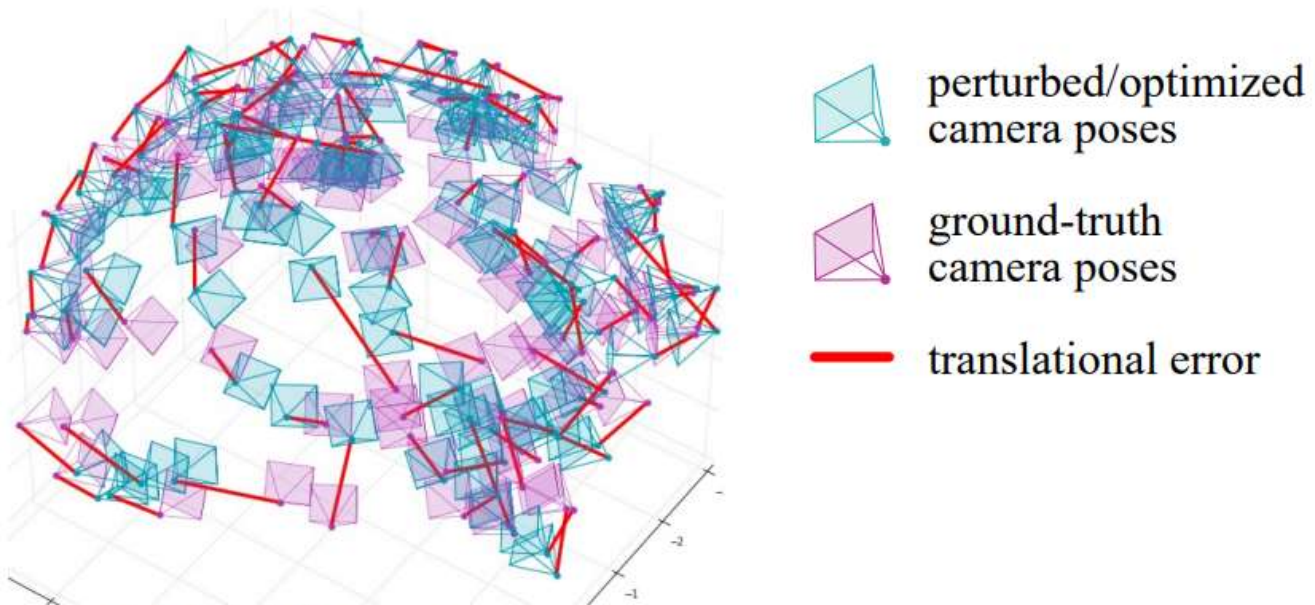
(b) initialization



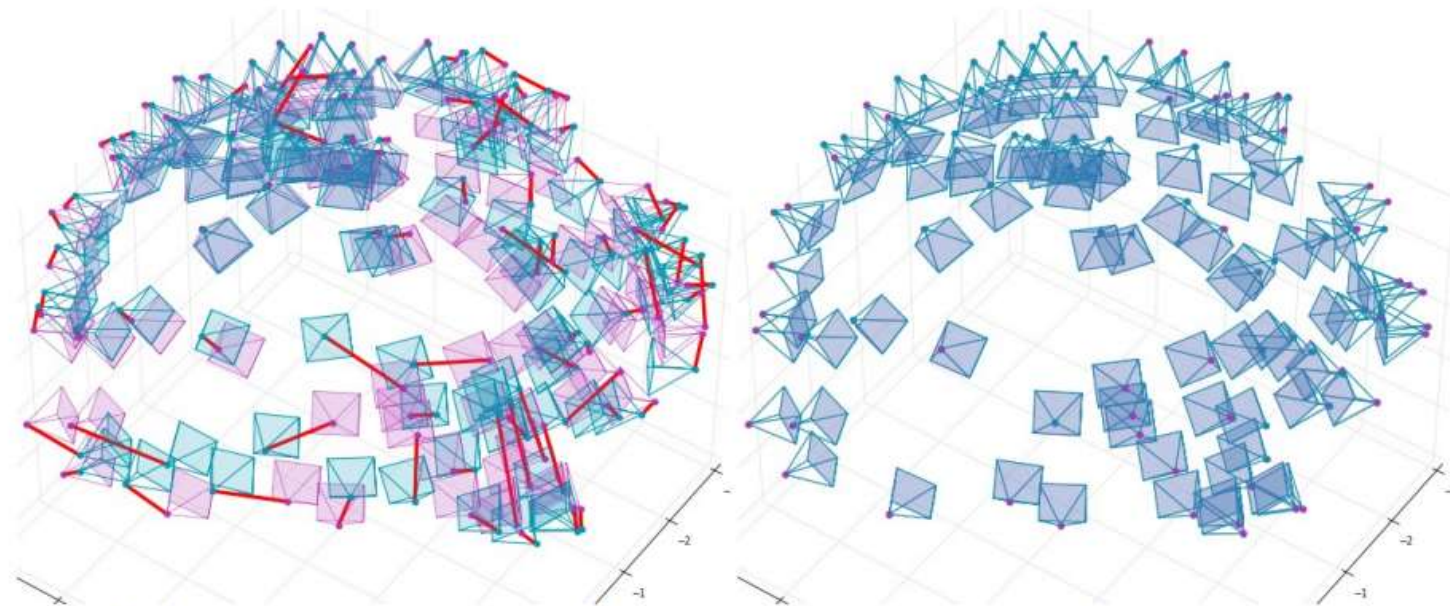
(c) ground-truth warps

Сравнение алгоритмов





(a) initial camera poses



(b) full positional encoding

(c) BARF (ours)

Здесь в искусственно
подготовленной сцене
немного подвигали камеры

ground truth



N/A



N/A



N/A

full pos. enc.



BARF (ours)



reference NeRF



image

depth

image

depth

image

depth

Источники

- Статья про BARF: <https://arxiv.org/abs/2104.06405>
- Статья про NeRF: <https://arxiv.org/abs/2003.08934>
- Документация COLMAP: <https://demuc.de/colmap/>
- Статья про SfM: <https://demuc.de/papers/schoenberger2016sfm.pdf>
- Статья на habr про SfM: <https://habr.com/ru/articles/301522/>