

Текст в 3D

Состав команды и куратор

- Николай
- Степан
- Антон (куратор)

Описание задачи

- Реализовать сервис, который из текстового описания объекта генерирует его 3D представление (mesh)
- Конвертировать полученное 3D представление в формат принимаемый популярными утилитами для работы с 3D объектами - Cinema4D, Blender

Краткий результат EDA

В результате исследования темы генерации text-to-3d, мы остановились на модели DreamFusion так как:

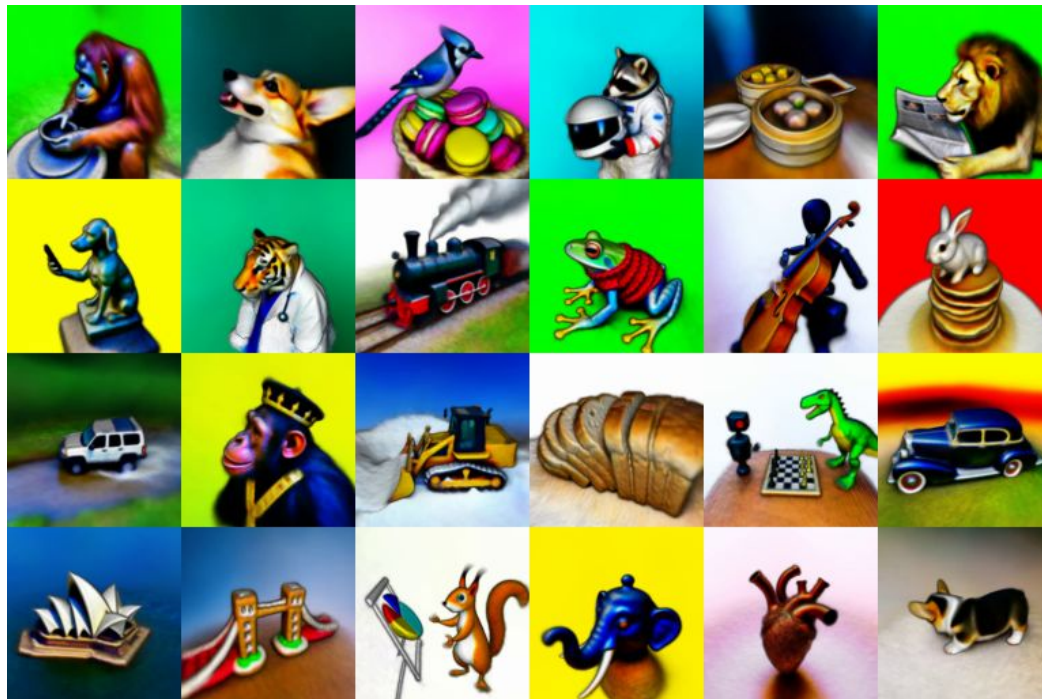
- Все последующие модели пользуются результатами DreamFusion, в первую очередь предложенной функцией потерь
- Dreamfusion не требует сбора данных, так как генерация 3D представления объекта основана на критике предобученной диффузионной модели
- Генерация 3D представлений, может быть, и хуже по сравнению с более новыми решениями, но хотелось бы разобраться с основными идеями и концепциями предложенным в DreamFusion

В качестве конкурирующих решений были рассмотрены Magic3d и Fantasia3d

Dreamfusion

Основными блоками
DreamFusion являются:

- NERF (Neural Radiance Fields)
- Score Distillation Sampling Loss



NeRF (Neural Radiance Fields)

Input Images



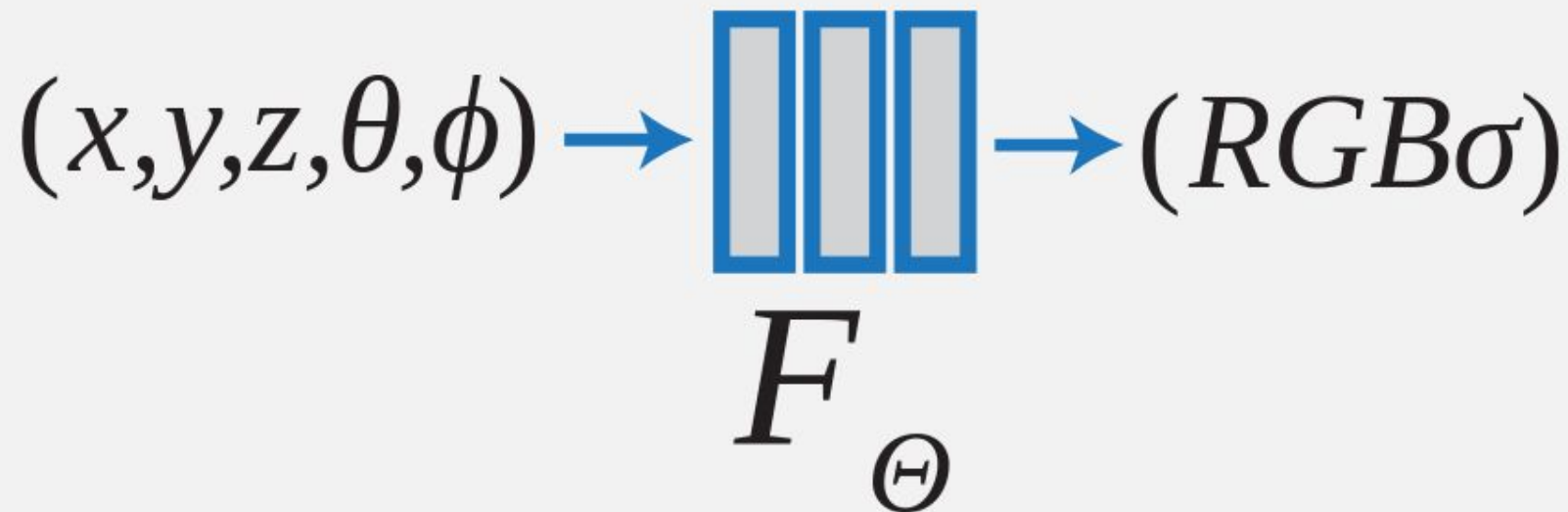
Optimize NeRF



Render new views



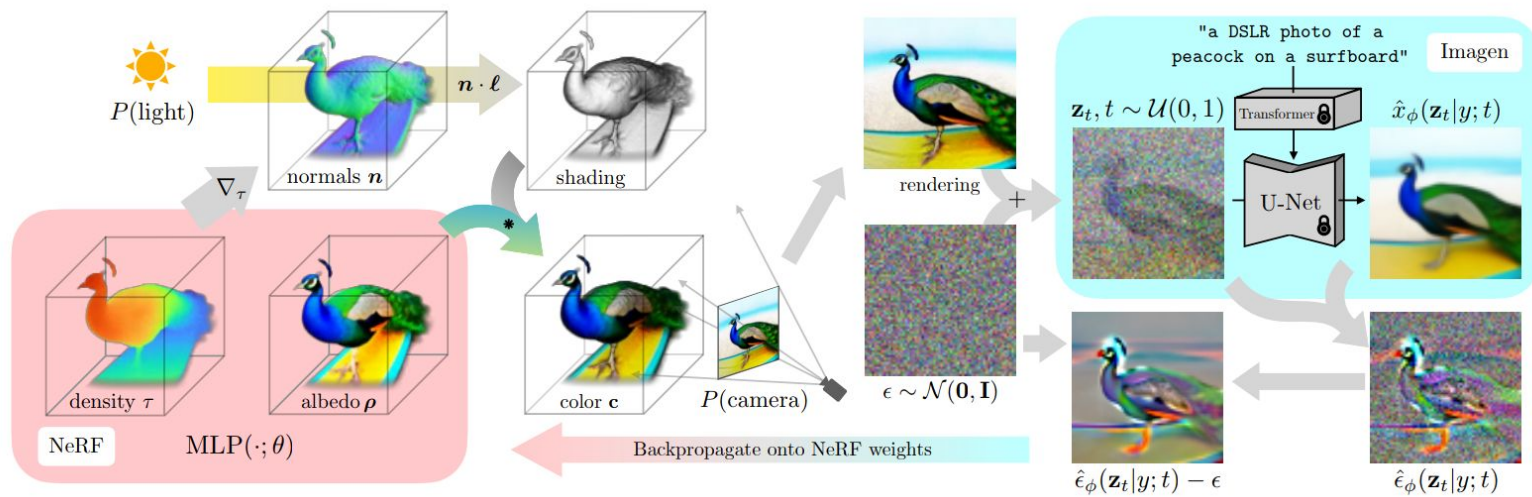
NERF (Neural Radiance Fields)



Score distillation sampling

$$\nabla_{\theta} \mathcal{L}_{\text{SDS}}(\phi, \mathbf{x} = g(\theta)) \triangleq \mathbb{E}_{t, \epsilon} \left[w(t) (\hat{\epsilon}_{\phi}(\mathbf{z}_t; y, t) - \epsilon) \frac{\partial \mathbf{x}}{\partial \theta} \right]$$

DreamFusion



Данные и их сбор

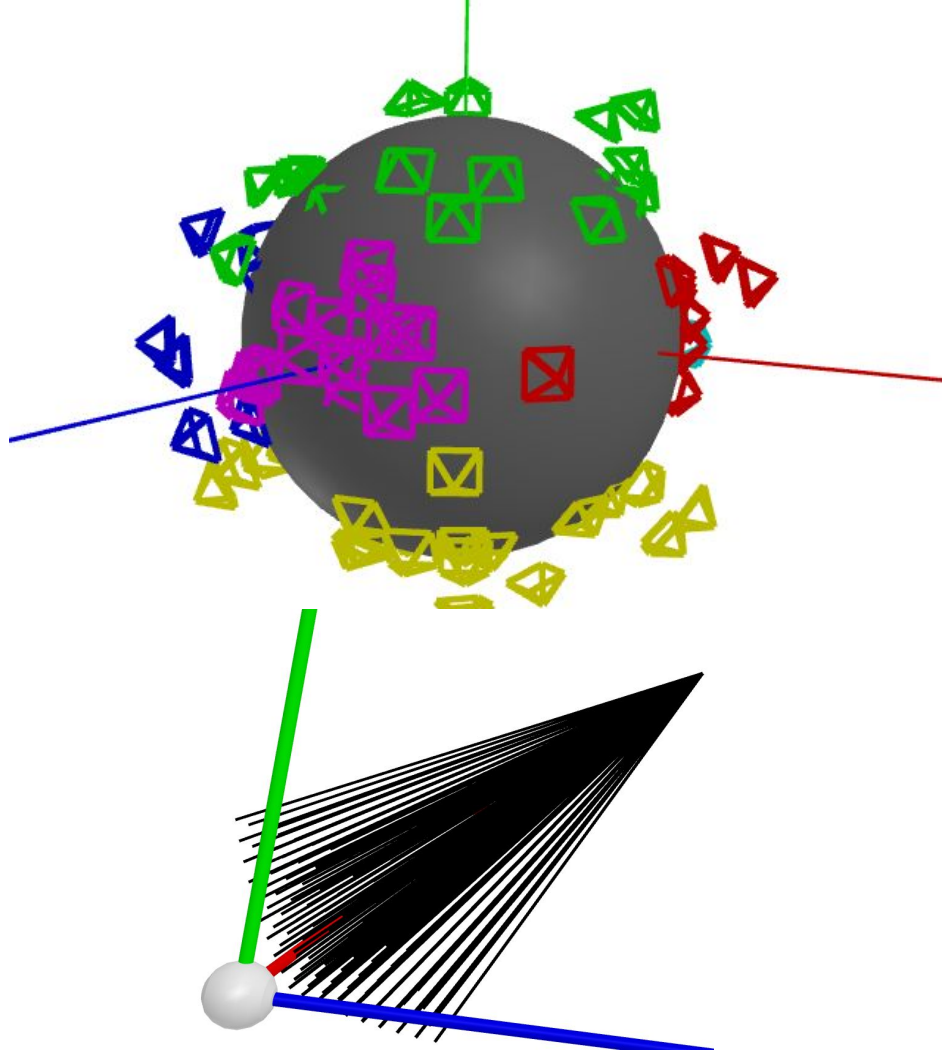
Для обучения NERF-а в алгоритме, предложенном DreamFusion, на вход подаются лучи опущенные в каждый пиксель с определенного ракурса камеры. NERF использует данные лучи для того, чтобы зарендерить 2D представление с выбранного ракурса камеры.

Задачей сбора данных в нашем проекте являлось придумать алгоритм генерации ракурсов камеры.

Данные и их сбор

Алгоритм заключался в следующем:

- Сгенерировать рандомные сферические координаты
- Перевести их в x, y, z
- Перевести x, y, z в координаты камеры $up, forward, right$
- Сгенерировать лучи с полученного ракурса камеры



Технические детали

- В качестве NERF мы взяли `tiny-nerf`
- В качестве диффузионного критика мы взяли `StableDiffusion 2.1`

Кто чем занимался

Николай - подготовка данных, реализация SDS Loss с помощью StableDiffusion.

Степан - рендер изображений с определенных ракурсов из NERF.

Результаты (so far)

- + Разобрались с pytorch
- + Разобрались с диффузионными моделями, NERF и SDS
- + У нас есть базовая реализация обучения
- + Есть бот
- + Обучение падает из-за нехватки памяти :)

Следующие шаги

- Довести до ума training loop
- Сделать CI/CD
- Положить все в контейнер

Ссылки

<https://arxiv.org/pdf/2209.14988.pdf> - DreamFusion

<https://arxiv.org/pdf/2003.08934v2.pdf> - NERF

<https://arxiv.org/pdf/2303.13873.pdf> - Fantasia3D

<https://arxiv.org/pdf/2211.10440.pdf> - Magic3D

<https://github.com/msalvaris/tiny-nerf> - Tiny-Nerf

<https://github.com/RichardTry/text-to-3d> - Наш Dreamfusion