

Синтез текстур с использованием диффузионных моделей

Мишан Алиев, Дима Баранчук, Кирилл Стуминский

Text-to-Image

- Подаем на вход промпт
- Возвращаем тензор $H \times W \times 3$



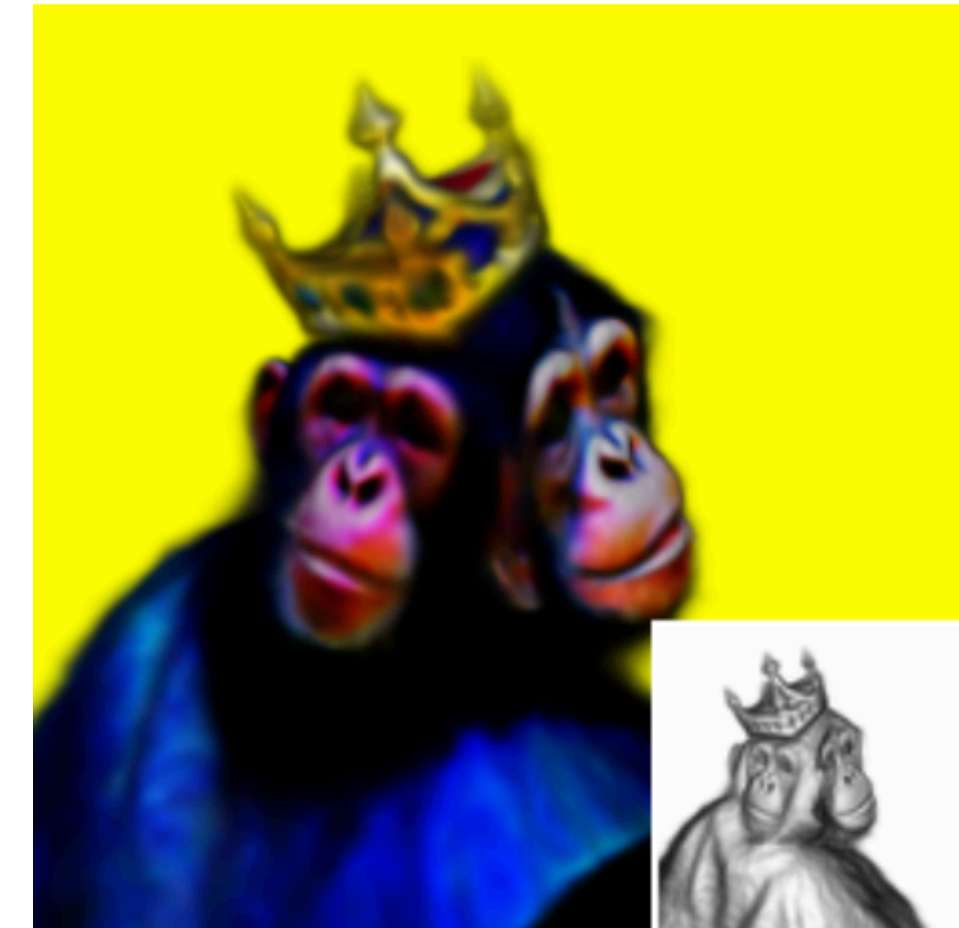
Text-to-3D

- Подаем на вход промпт
- Учим объемное 3D представление
 - Поле плотности $\sigma : \mathbb{R}^3 \rightarrow \mathbb{R}_+$
 - Светимость $C : \mathbb{R}^3 \times S^2 \rightarrow \mathbb{R}^3$
- Неявные представления / оплаты



Минусы

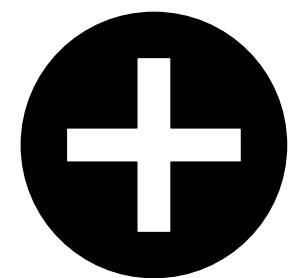
- Сложная непредсказуемая геометрия
- Высокий реализм, но статические сцены
 - Статическая геометрия
 - Статическое освещение
- Отрыв от инструментов 3D графики



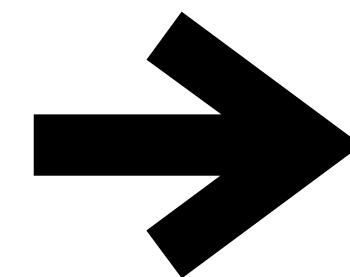
Синтез текстур



Фиксированный меш



Текстовый запрос



Текстурированный меш

Как передать цвет?

- Просто
 - RGB в каждой точке $C : [0,1]^2 \rightarrow [0,1]^3$
- Предельно сложно
 - Зависимость от точки обзора и времени

$$C : [0,1]^2 \times S^2 \times T \rightarrow [0,1]^3$$

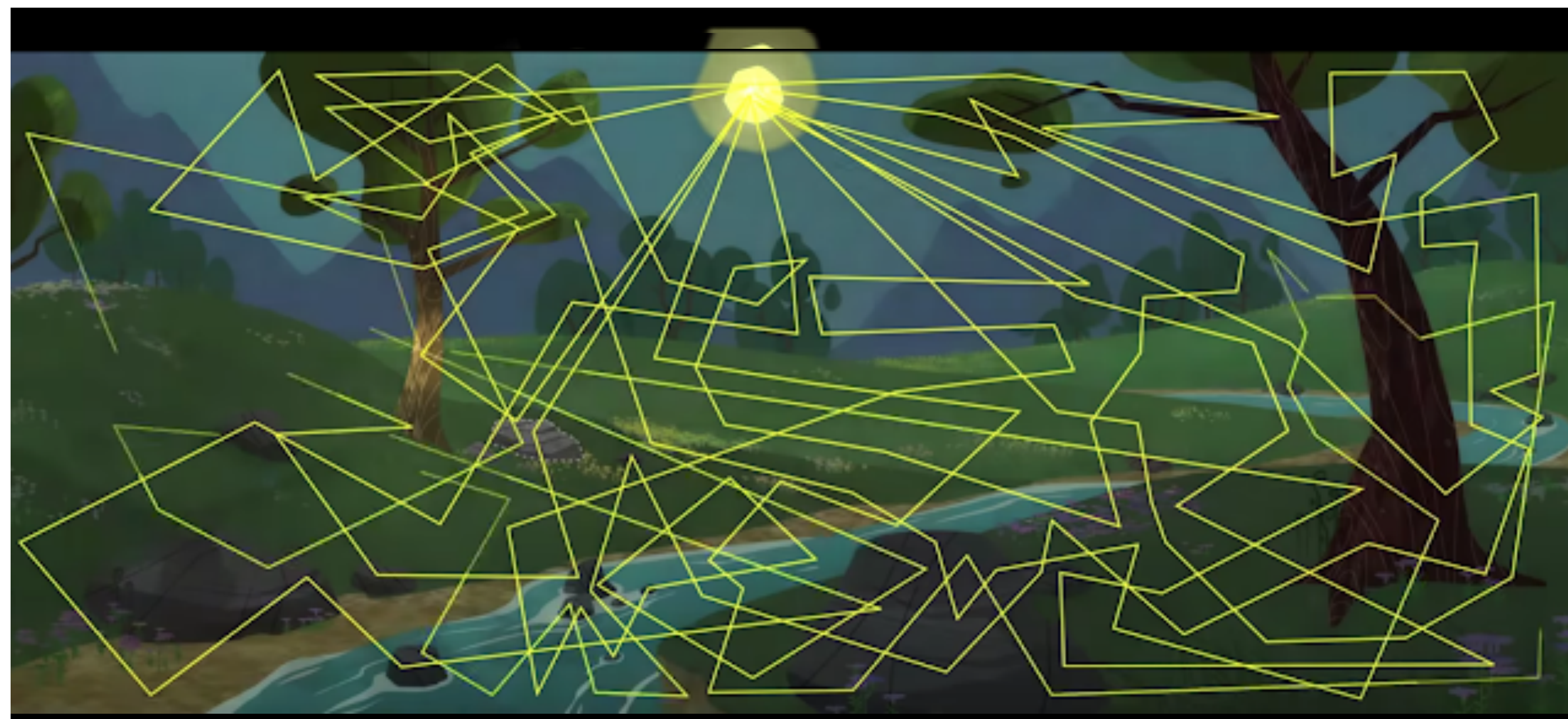
- Есть ли более эффективные представления?



**Немножко о том как
формируется изображение**

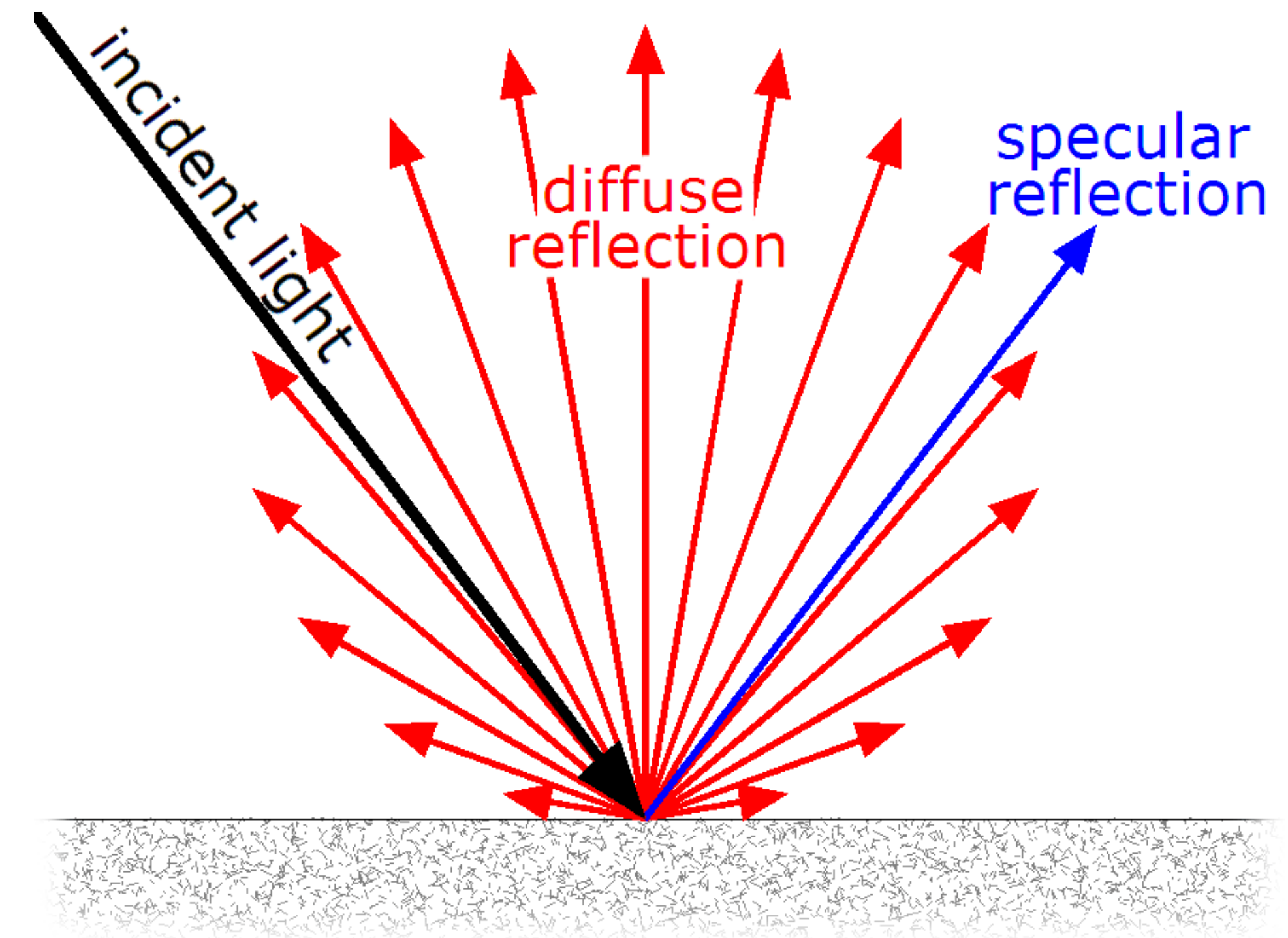
Physically Based Rendering (PBR)

- Идея: симуляция распространения света
- Выпускаем множество лучей
- Лучи отражаются о поверхности сцены
- Регистрируем те лучи, которые в конце концов дошли до сенсора камеры



Взаимодействие света

- Диффузное рассеяние
 - Направления отражения равновероятны
- Зеркальное отражение
 - Угол отражения равен углу преломления
- В общем случае: BRDF
 - Bidirectional Reflectance **Distribution** Function
 - Функция $f_r(\omega_i, \omega_r) : [0,1]^2 \times [0,1]^2 \rightarrow \mathbb{R}$



Как это всё работает на практике?

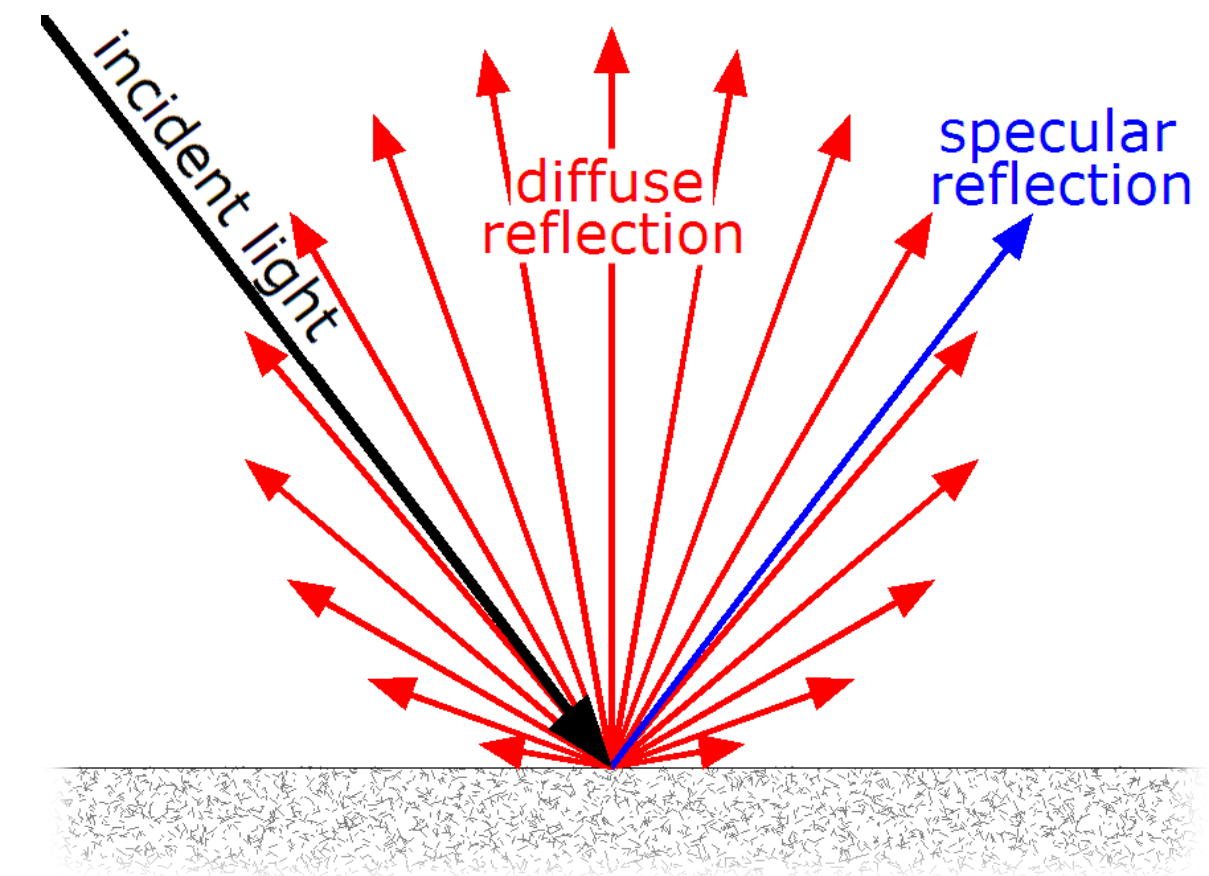
- Запускать симуляцию долго и дорого, используют приближения (например, растеризация)
- Напрямую представлять BRDF функцию очень сложно, в индустрии распространен стандарт от студии Disney



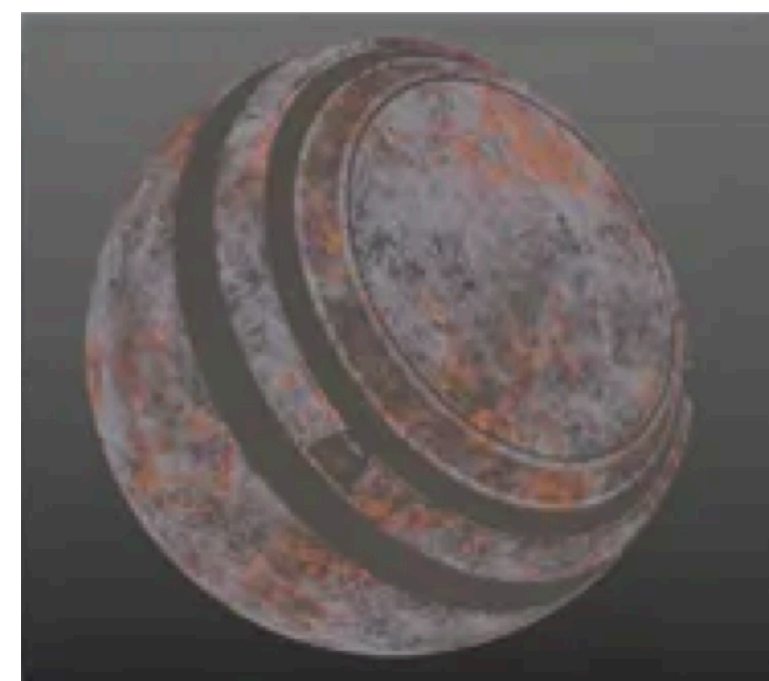
PBR текстуры

Материалы, состоящие из

- диффузной карты (обычный цвет) k_c
- металличности k_m
- шероховатости k_r
- карты нормалей k_n

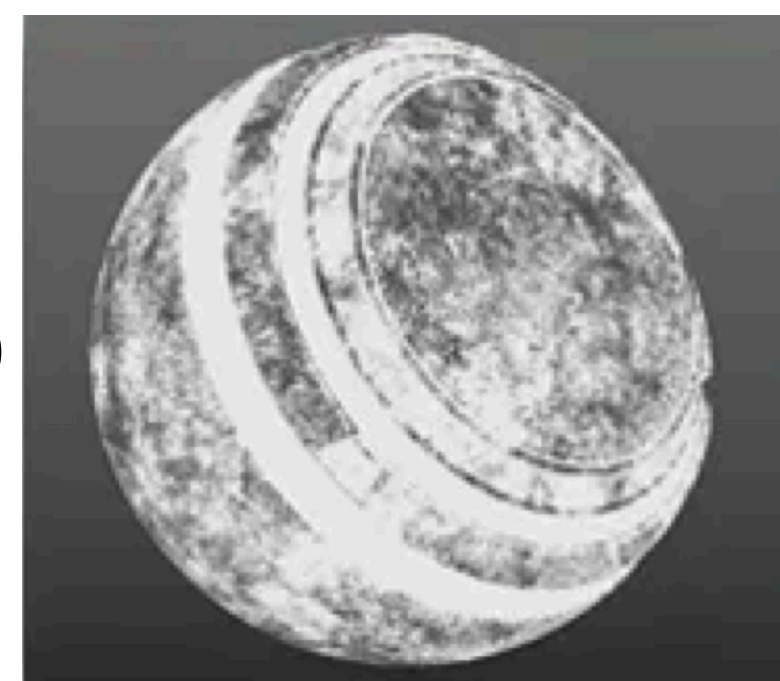


Изображение



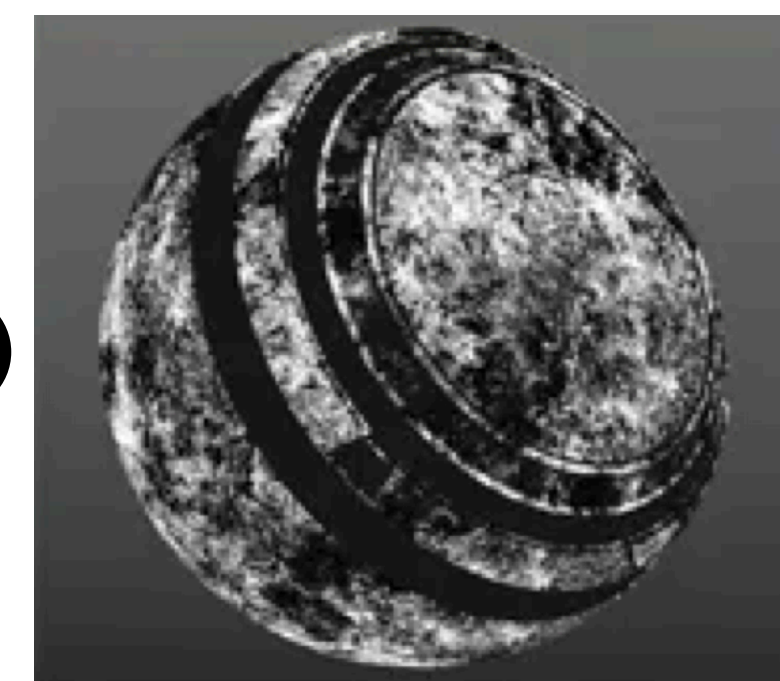
Диффузная карта

+



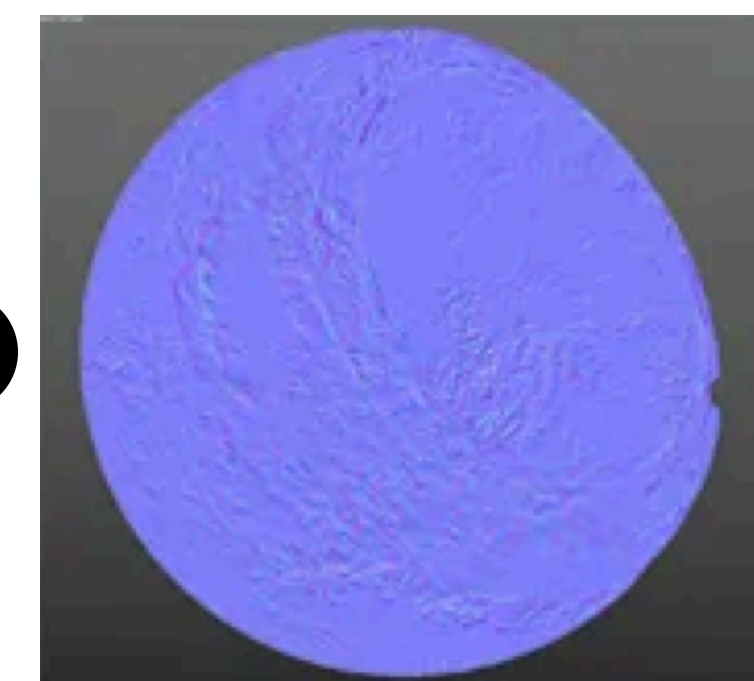
Шероховатость

+



Металличность

+



Карта нормалей



Просто цвет

Vs.

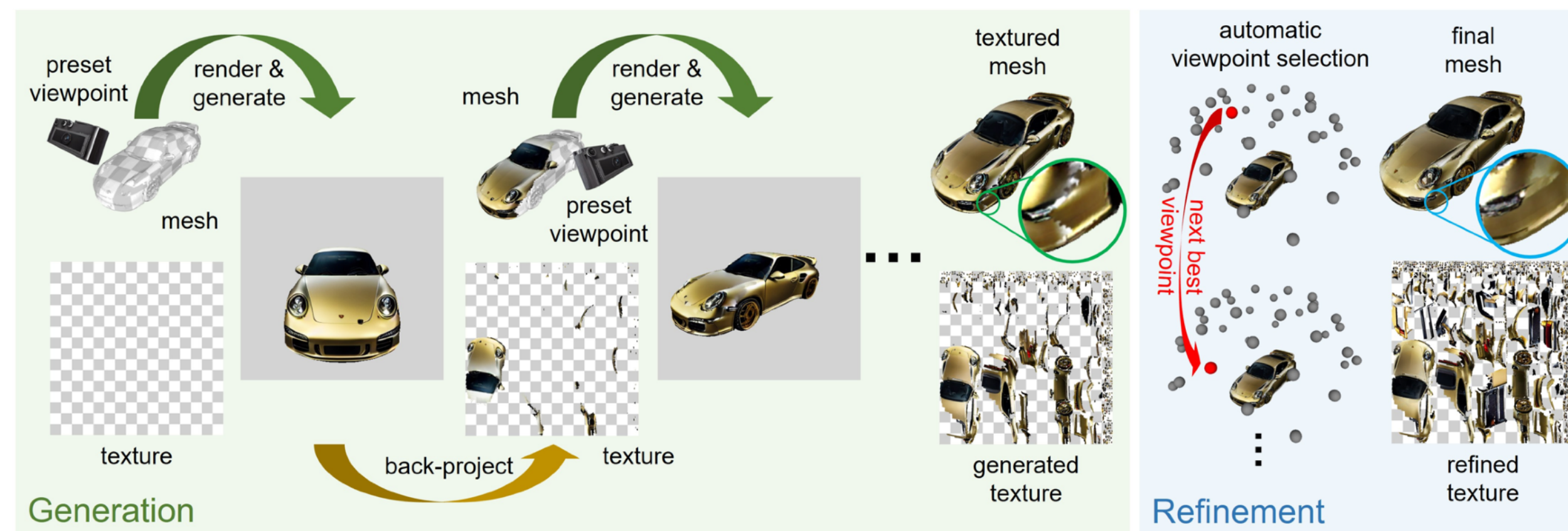


Не просто цвет

Генеративные модели

Так как одеть 3D-модель?

- Предобученные модели знают как выглядит окружающий мир
- Не PBR-подходы: сгенерировать ракурс, а потом спроецировать цвета

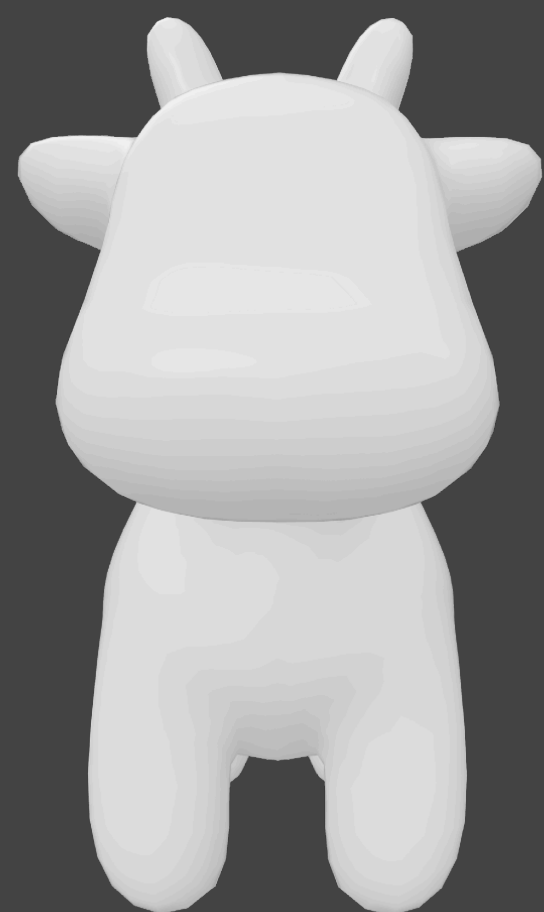


- Наш путь: PBR-подходы на основе Score Distillation Sampling



Как работает наш метод

Фиксированный меш



Случайный свет



Случайная
позиция камеры

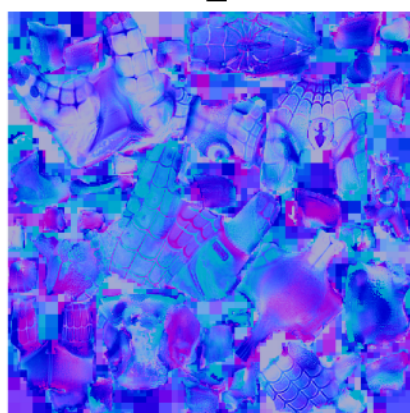


Обучаемые параметры текстуры

diffuse_texture



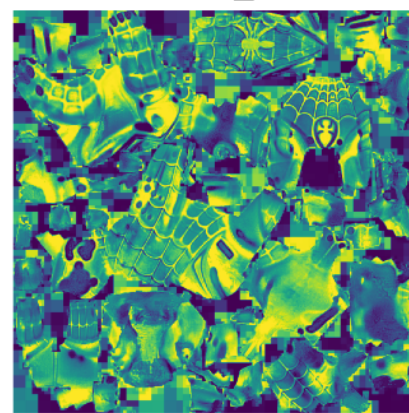
normals_texture



metallic_texture



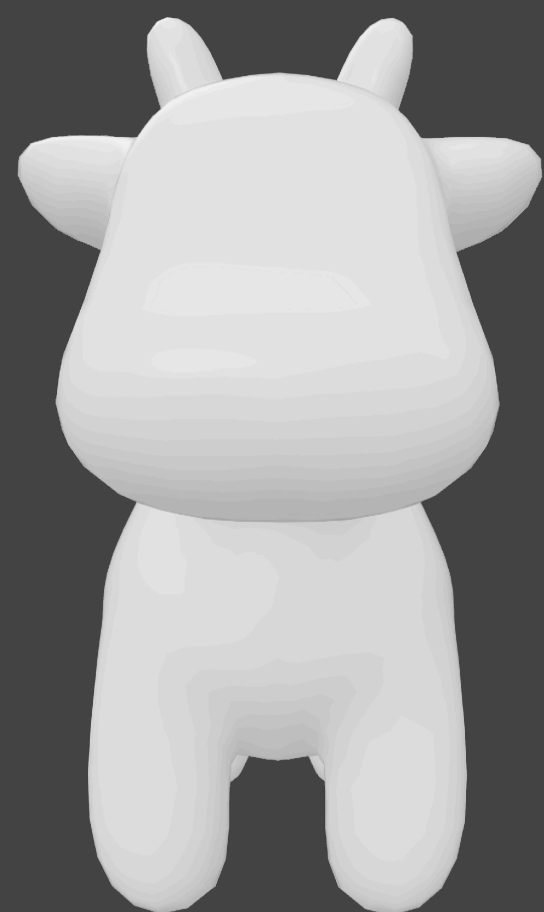
roughness_texture





Как работает наш метод

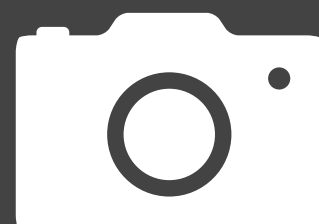
Фиксированный меш



Случайный свет



Случайная
позиция камеры

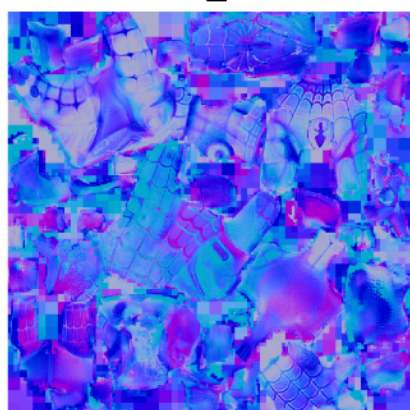


Обучаемые параметры текстуры

diffuse_texture



normals_texture



metallic_texture



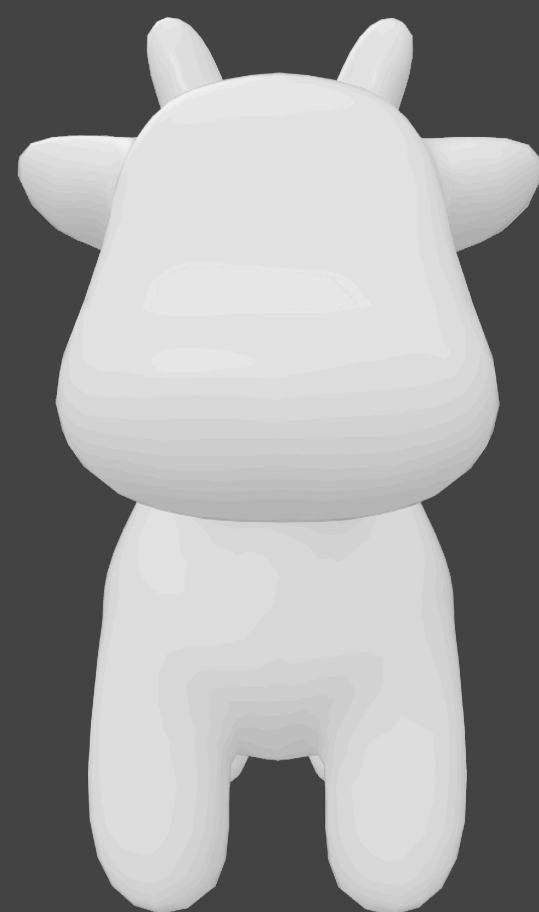
roughness_texture





Как работает наш метод

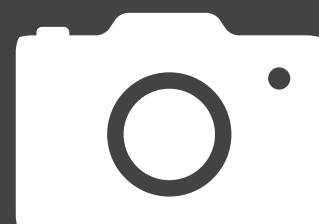
Фиксированный меш



Случайный свет



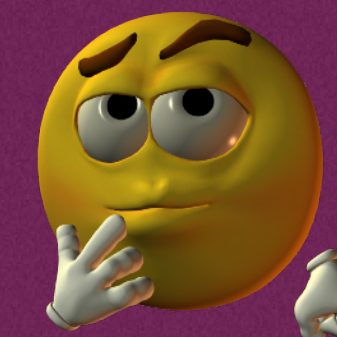
Случайная
позиция камеры



Нейросеть с
насмотренностью



«А насколько эта
2D картинка
правдоподобна?»»

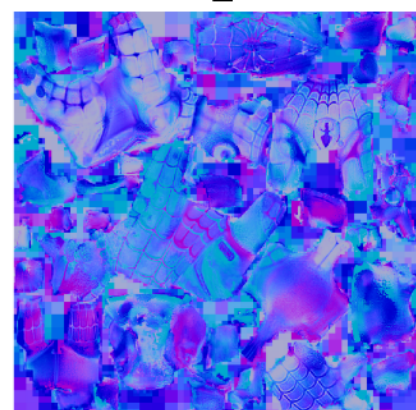


Обучаемые параметры текстуры

diffuse_texture



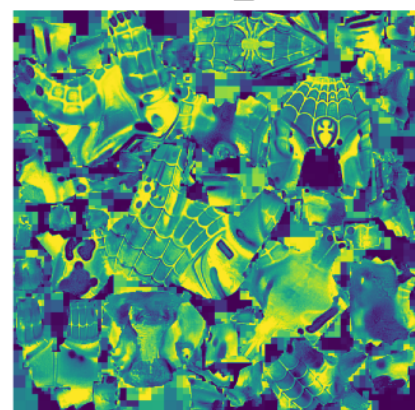
normals_texture



metallic_texture




roughness_texture






Как работает наш метод

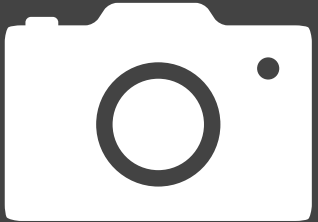
Фиксированный меш



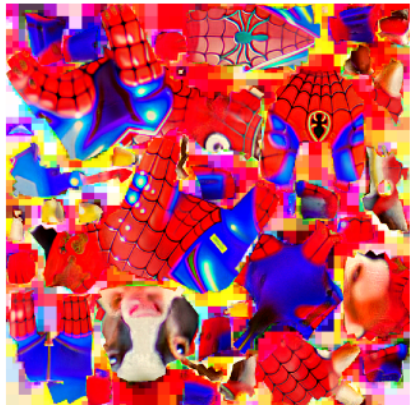
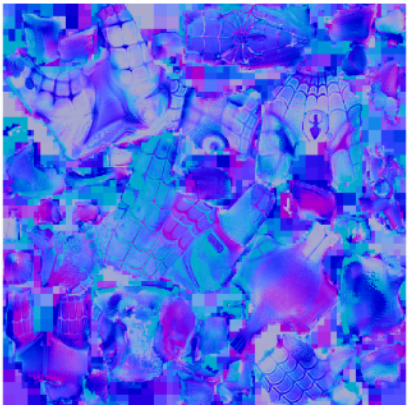
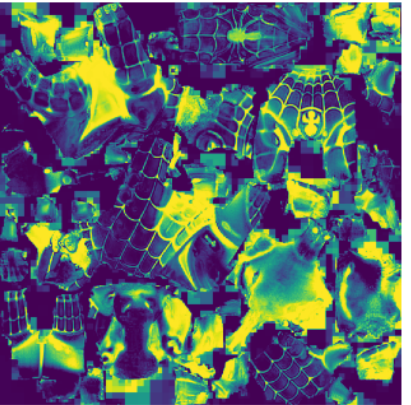
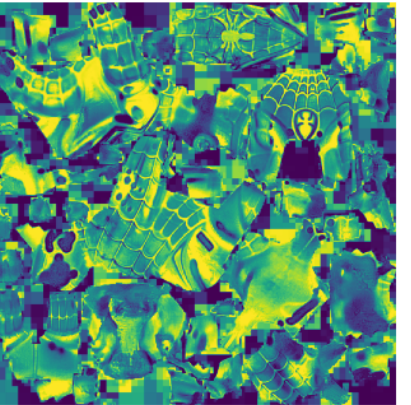
Случайный свет



Случайная позиция камеры




Обучаемые параметры текстуры


diffuse_texture	normals_texture	metallic_texture	roughness_texture
			



Нейросеть с насмотренностью



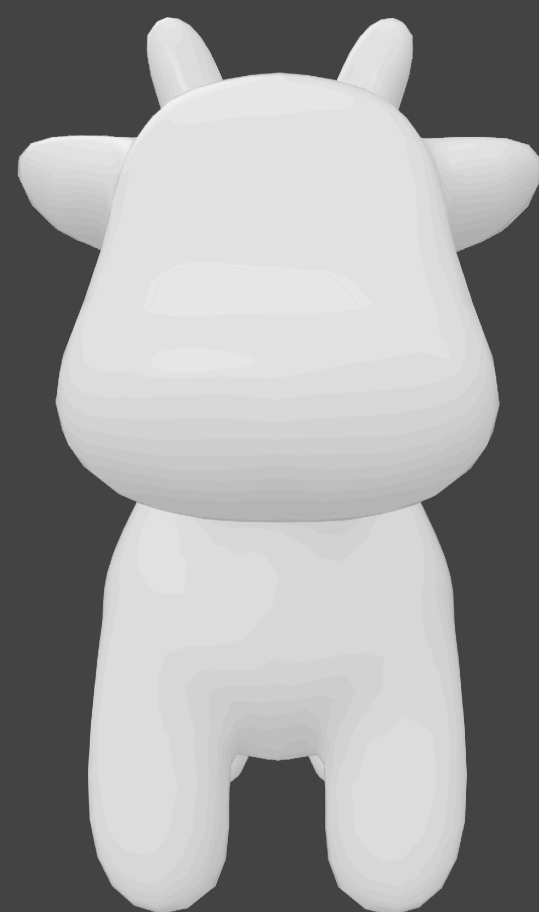
«А насколько эта 2D картинка правдоподобна?»»






Как работает наш метод

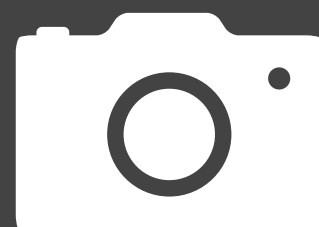
Фиксированный меш



Случайный свет



Случайная
позиция камеры

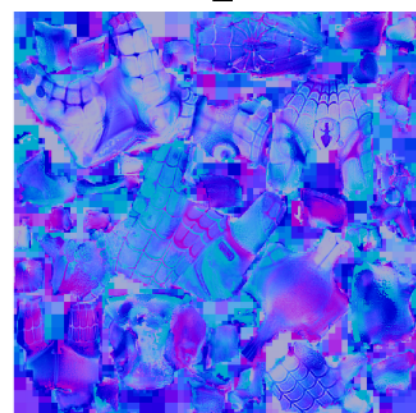


Обучаемые параметры текстуры

diffuse_texture



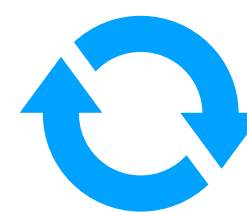
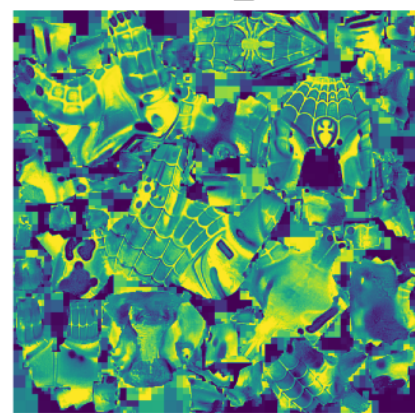
normals_texture



metallic_texture



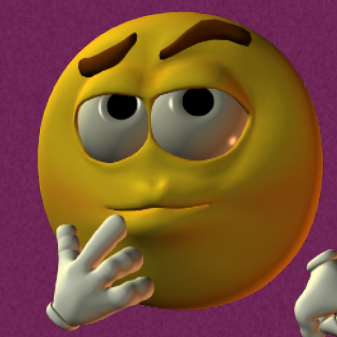
roughness_texture



Нейросеть с
насмотренностью



«А насколько эта
2D картинка
правдоподобна?»»

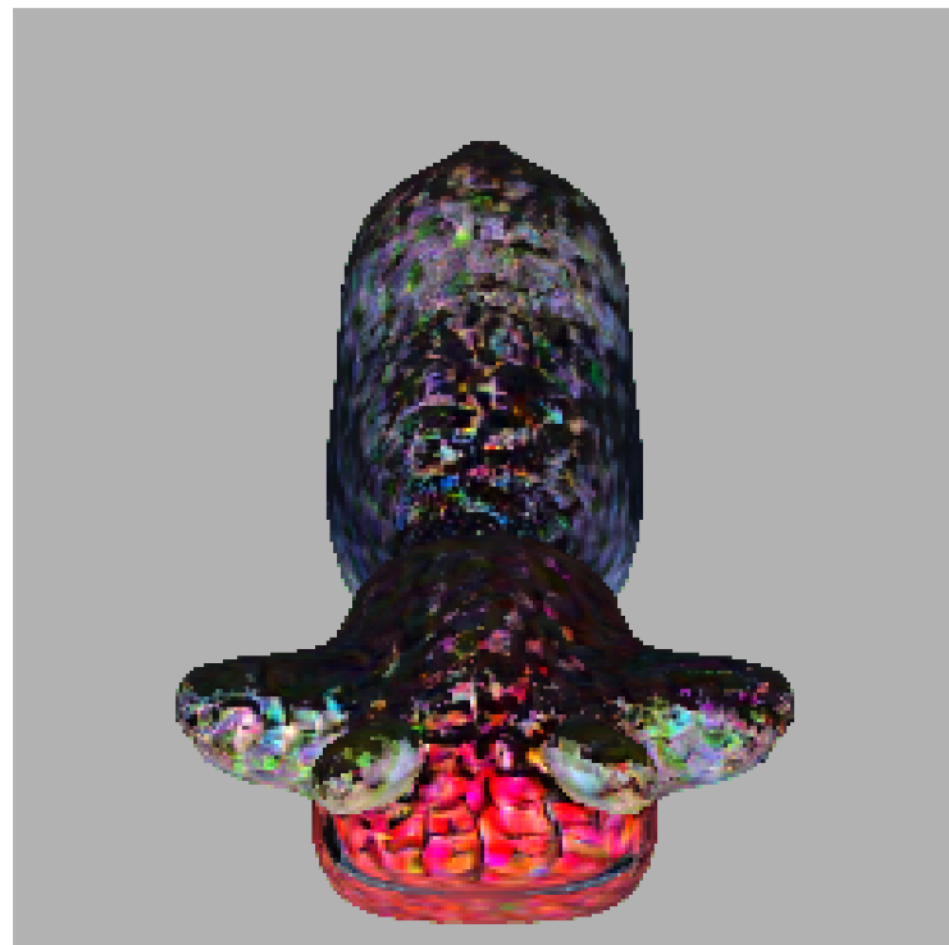


Что под капотом?

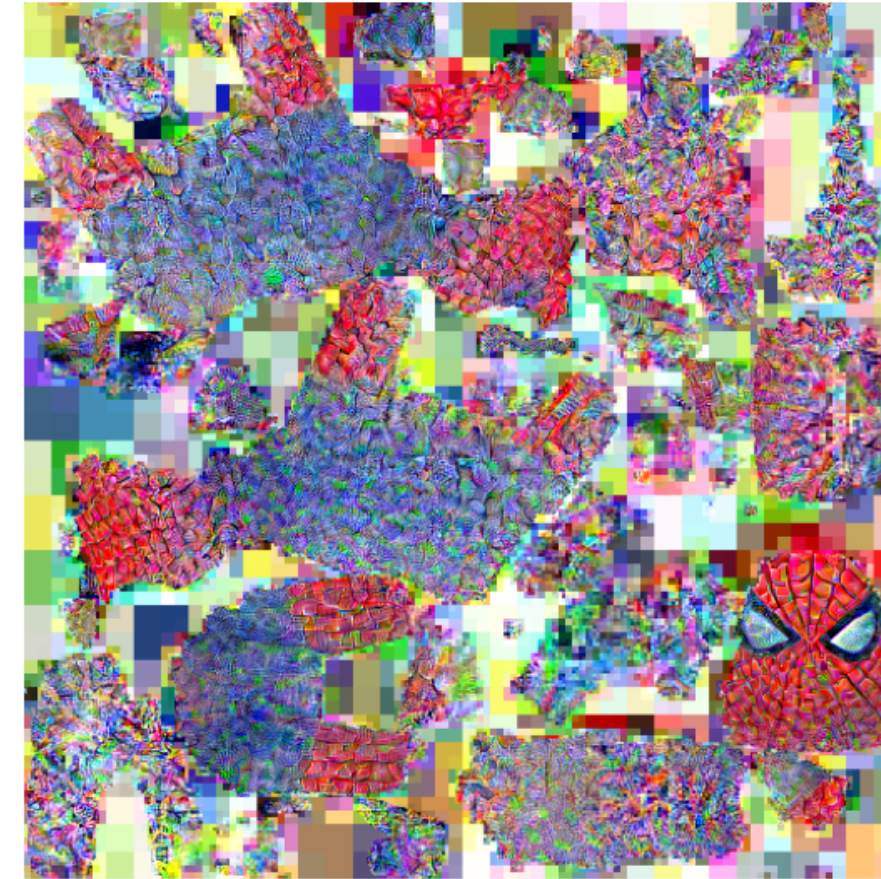
- Нарисовали картинку $x_0 = g(\theta, z)$ и зашумили $x_t = \alpha_t x_0 + \sigma_t \epsilon$
- С помощью диффузионной модели по x_t оцениваем, как выглядела исходная картинка
- По оценке \hat{x}_0^t определяем градиент как

$$\nabla_{\theta} \mathcal{L}_{\text{SDS}}(\theta) = \mathbb{E}_{t, \epsilon, c} \left[\frac{\omega(t)}{\gamma(t)} (\mathbf{x}_0 - \hat{\mathbf{x}}_0^t) \frac{\partial g(\theta, c)}{\partial \theta} \right]$$

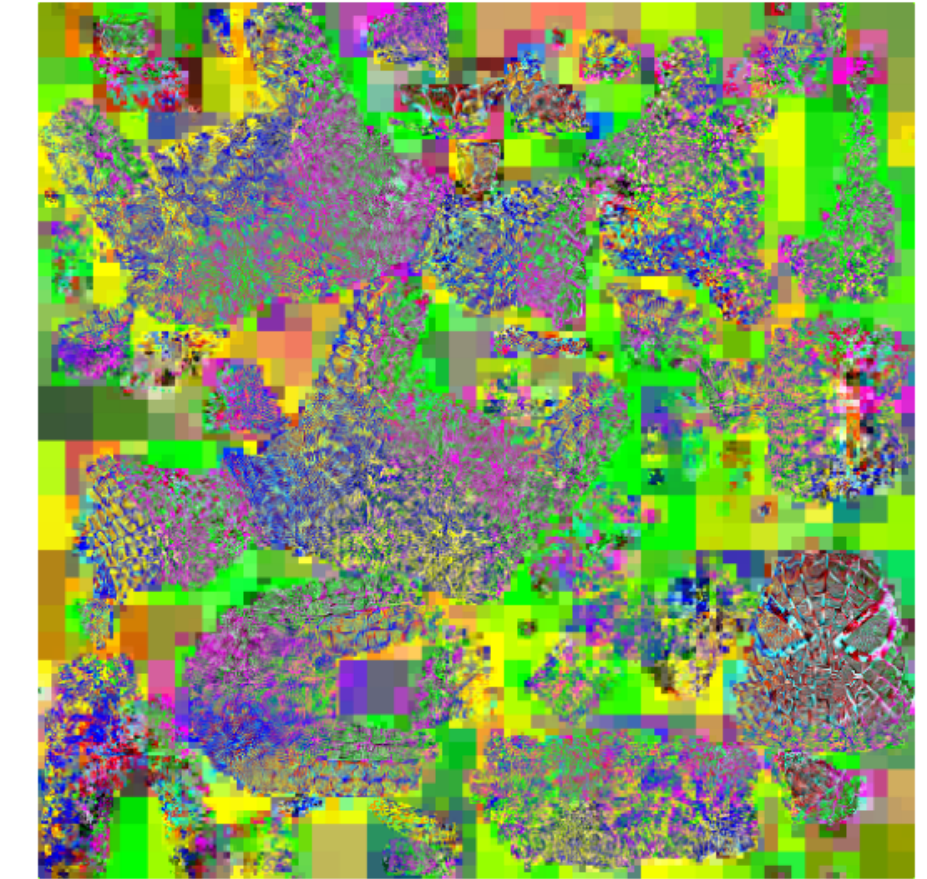
Опыт 1: латентные модели для текстур



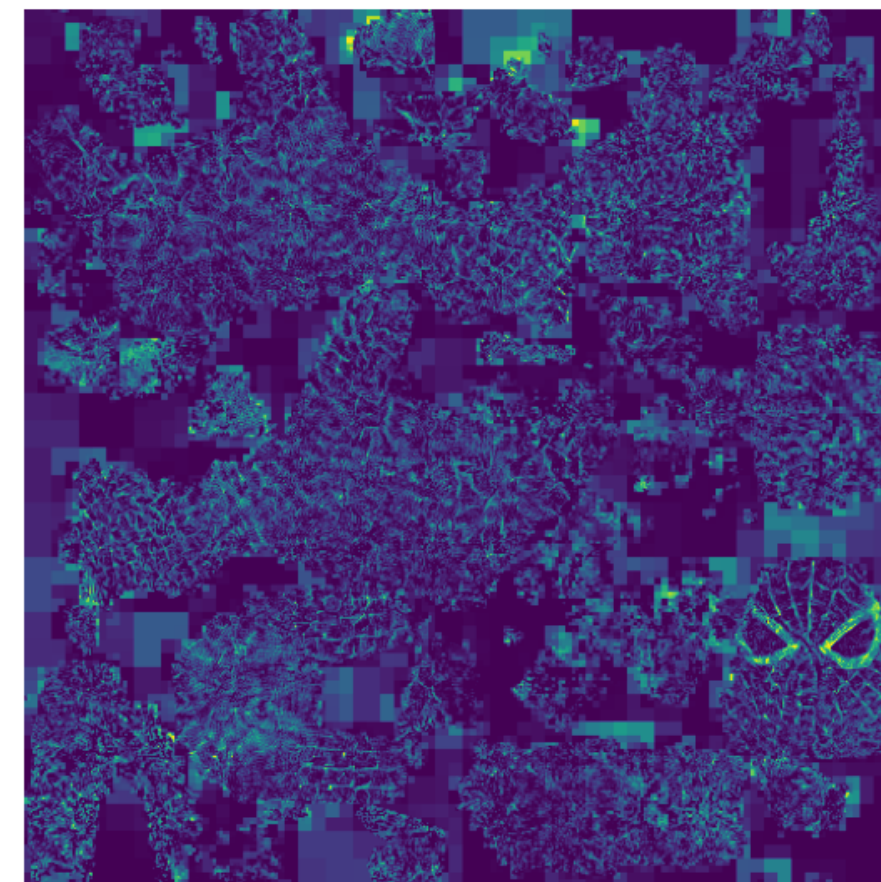
Albedo



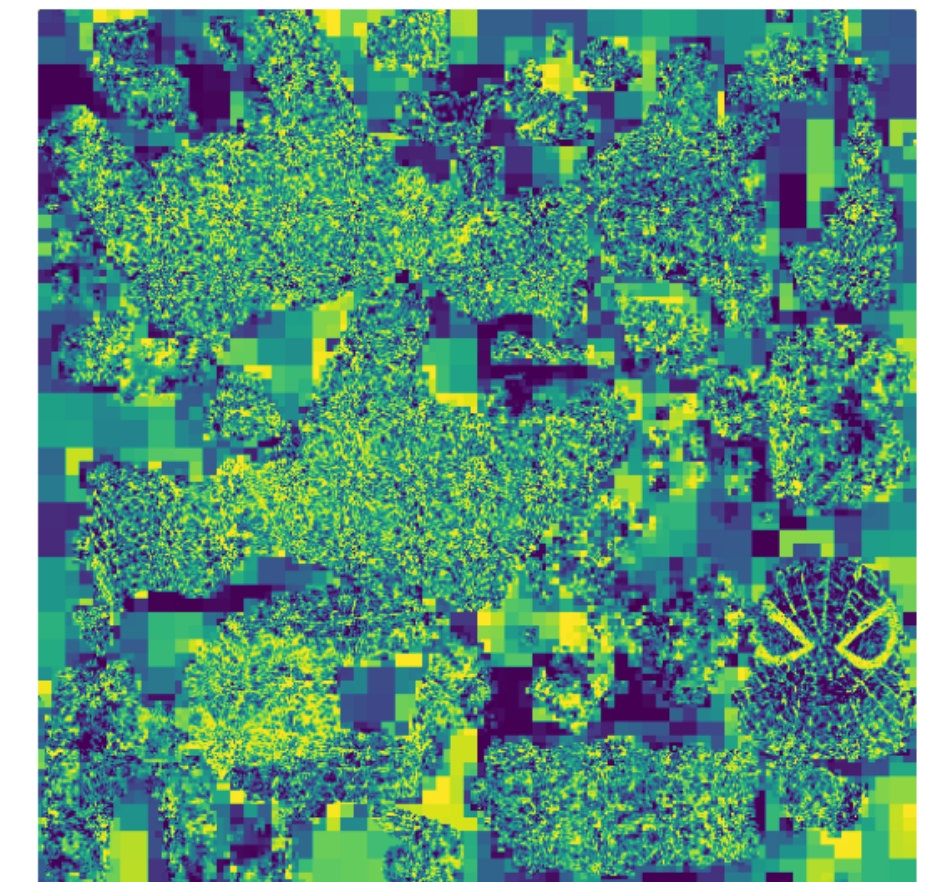
Normals



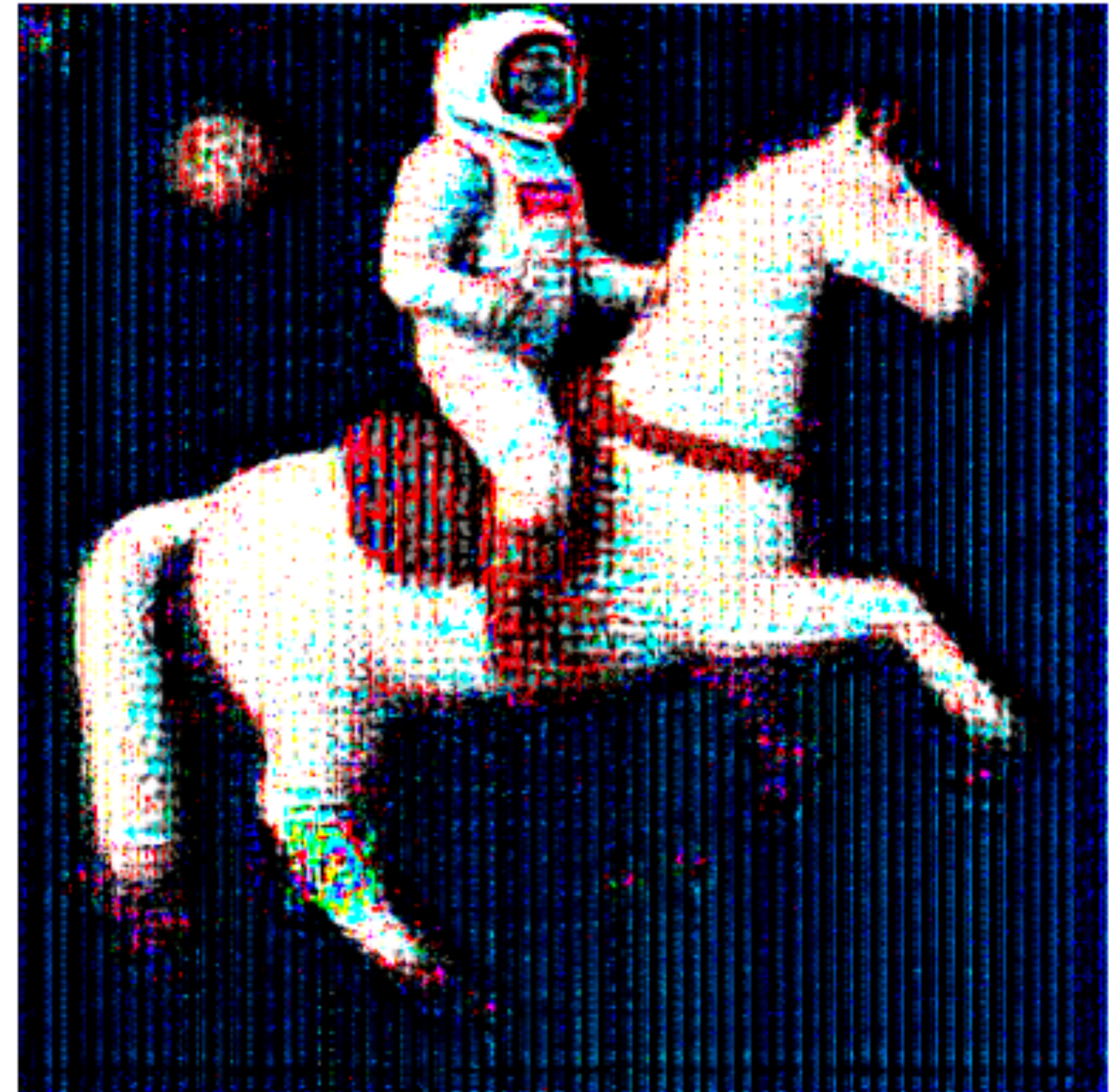
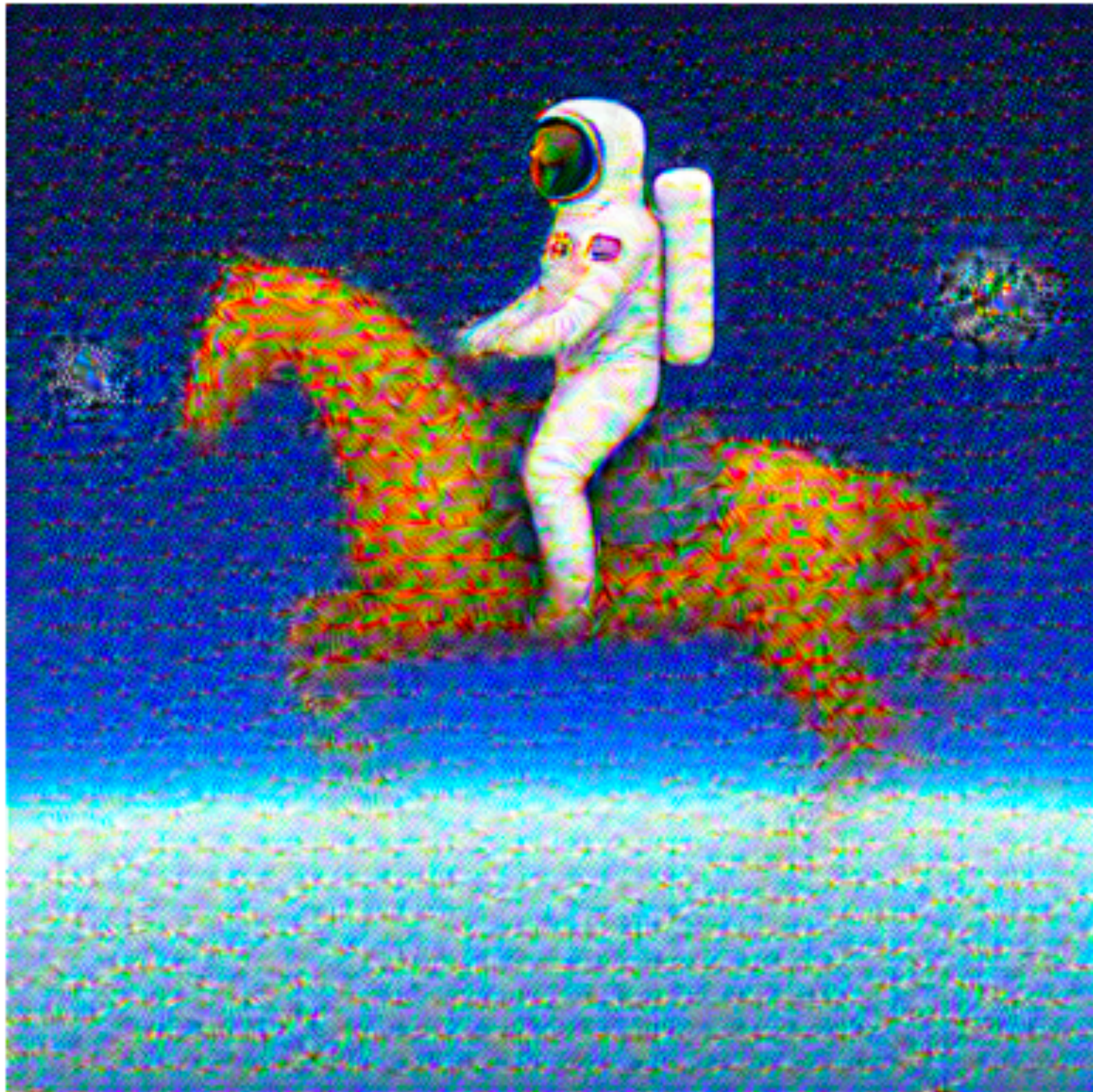
Metallic



Roughness



SDS с латентными моделями для картинок





Решение через регуляризацию

- Мы представляем текстуры напрямую как несколько изображений
- Paint-It (CVPR2024) предлагает Deep Image Prior параметризацию

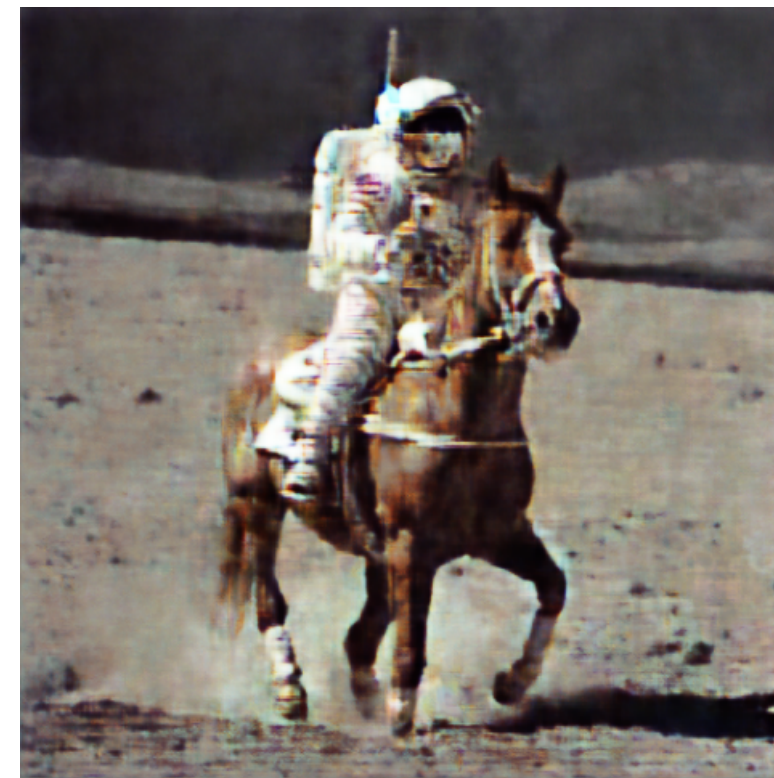
Original



RGB, latent L_2
PSNR=16.67



DIP, latent L_2
PSNR=22.53



RGB, rgb L_2
PSNR=59.58



DIP, rgb L_2
PSNR=36.85



DIP-параметризация для каскадных моделей



Опыт 2: усложнения SDS

Dreamfusion

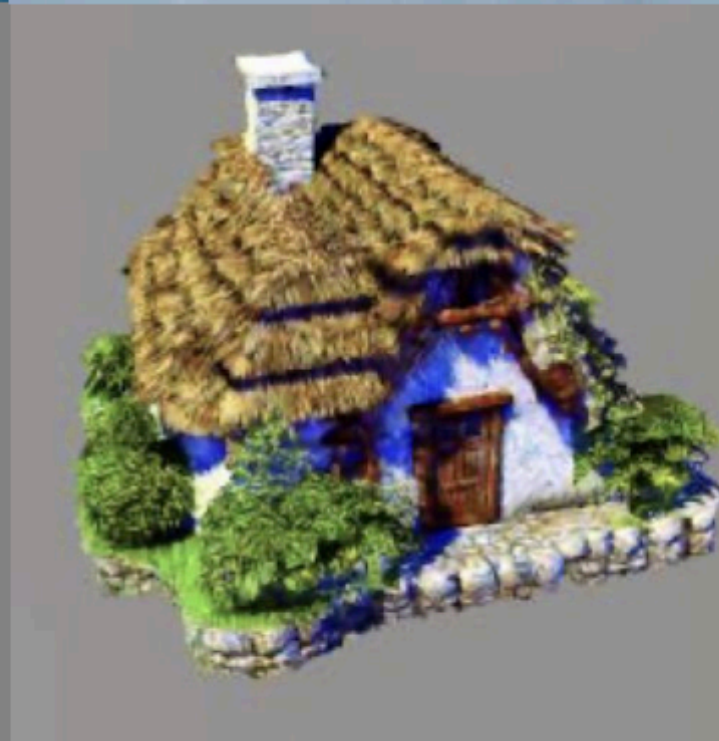
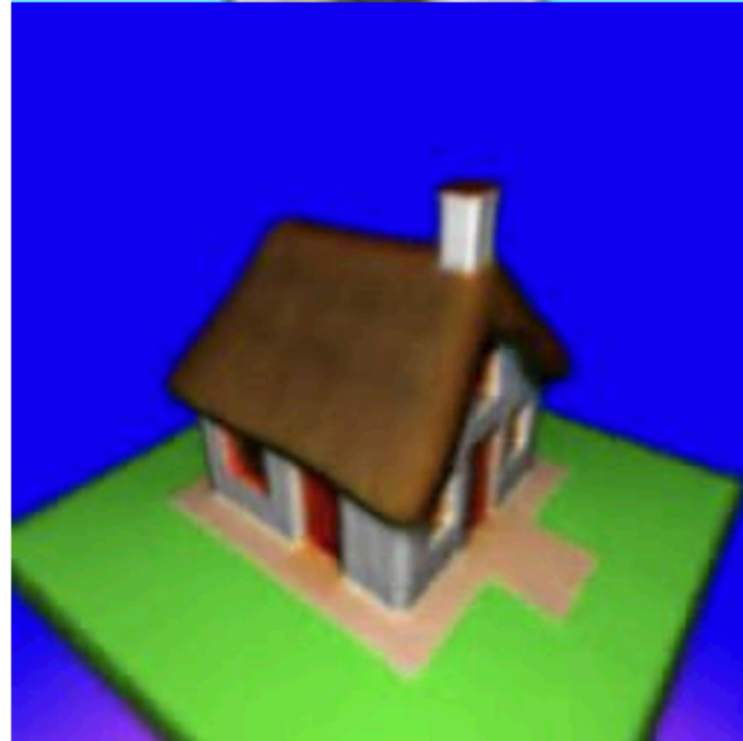
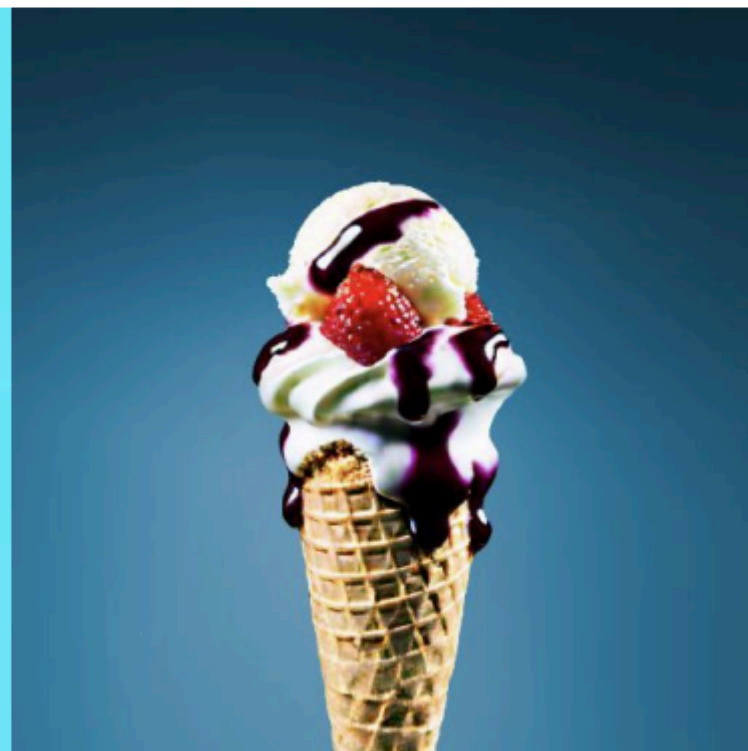
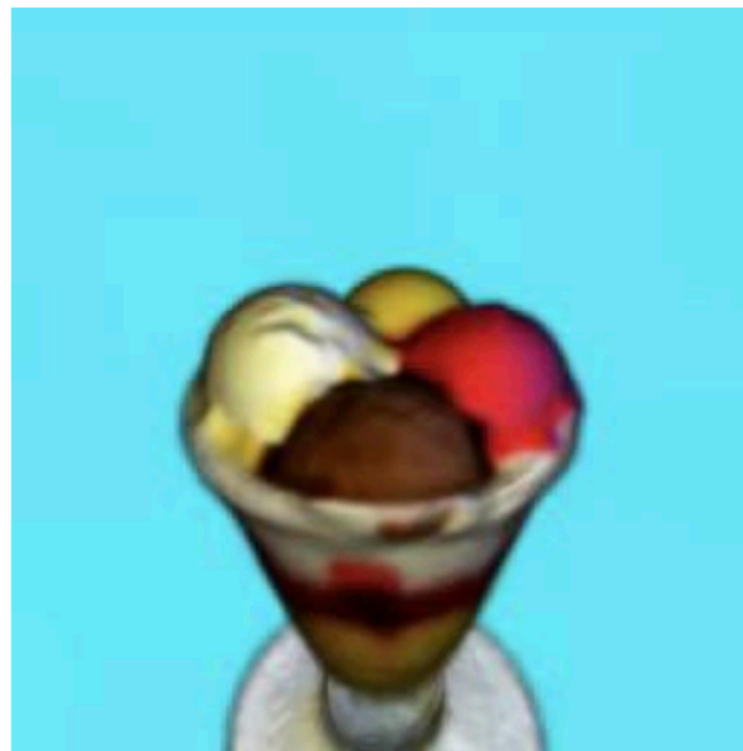
NFSD

ProlificDreamer

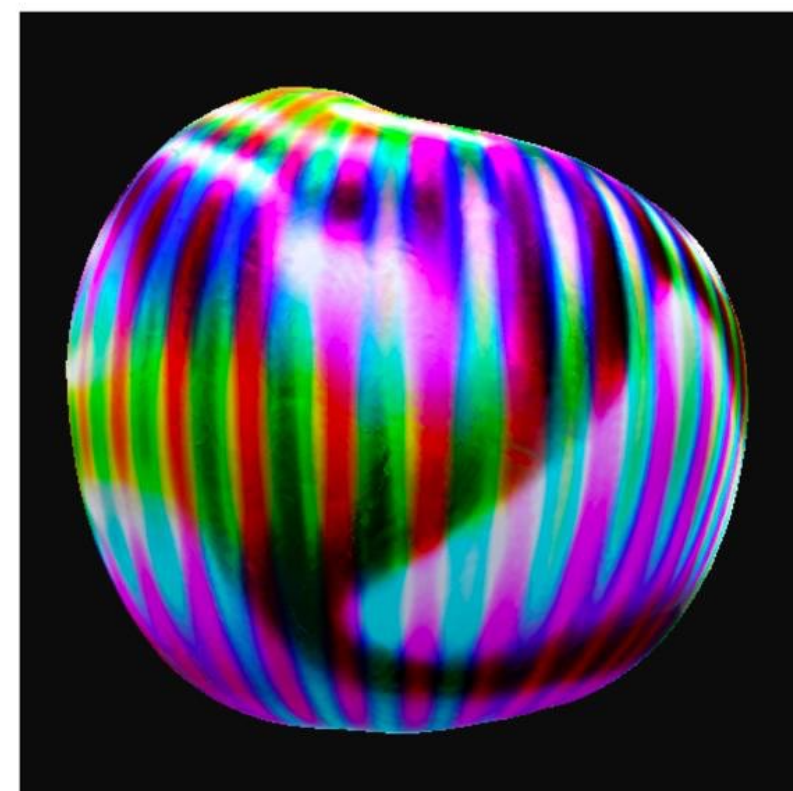
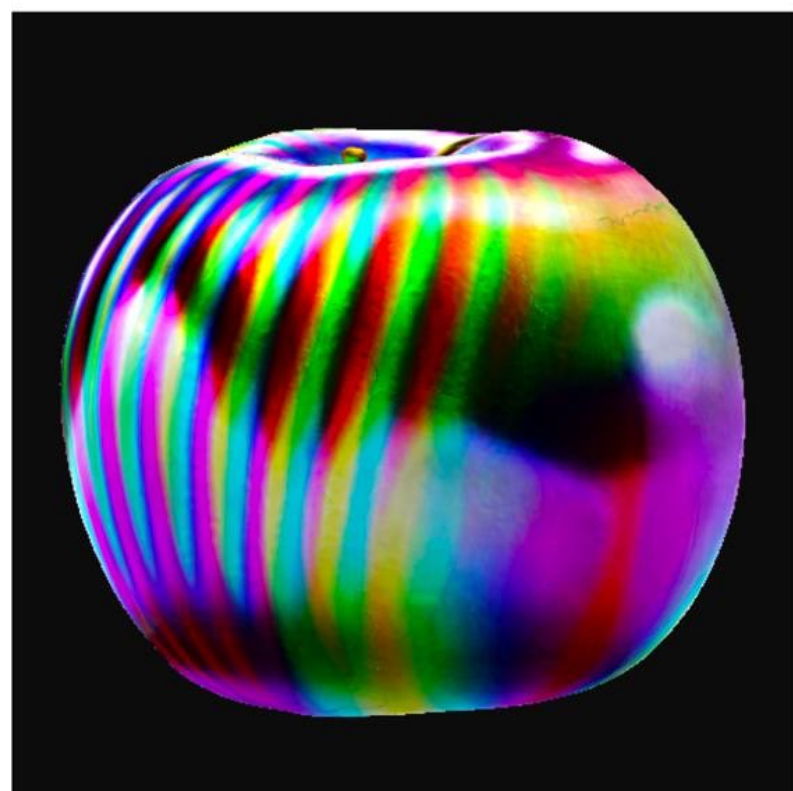
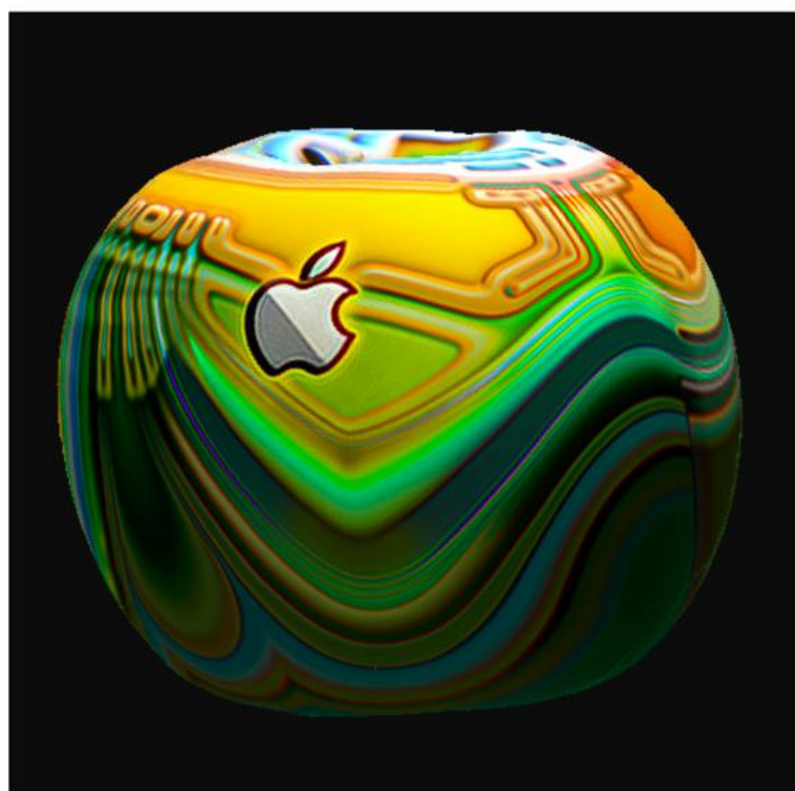
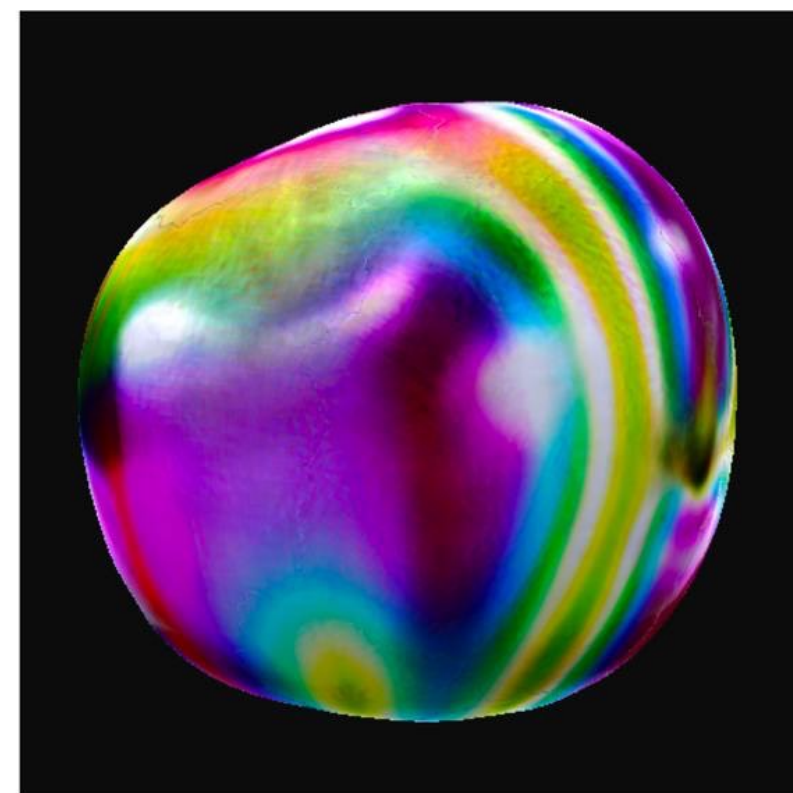
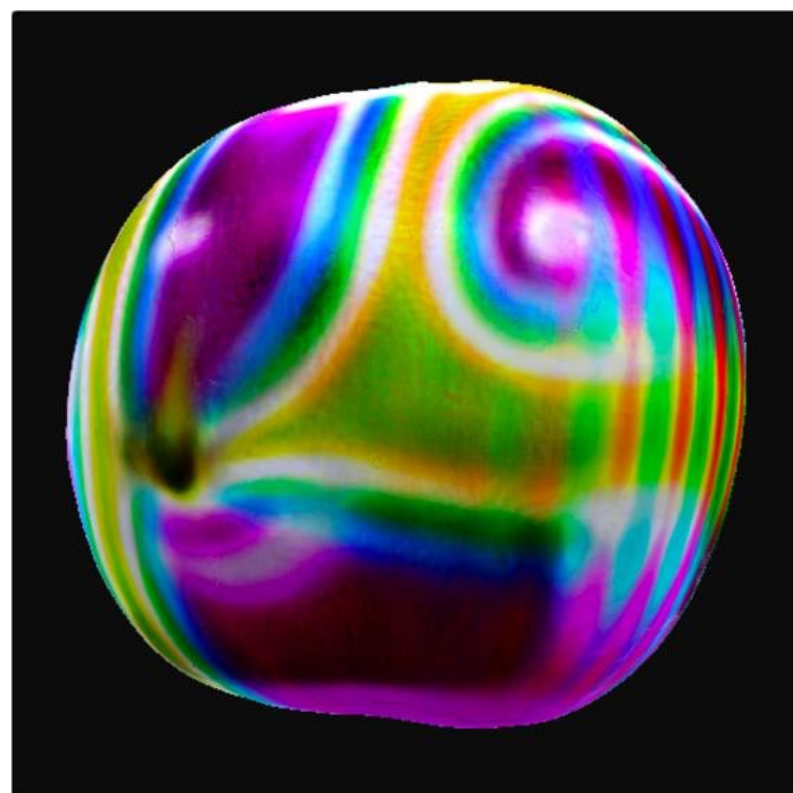
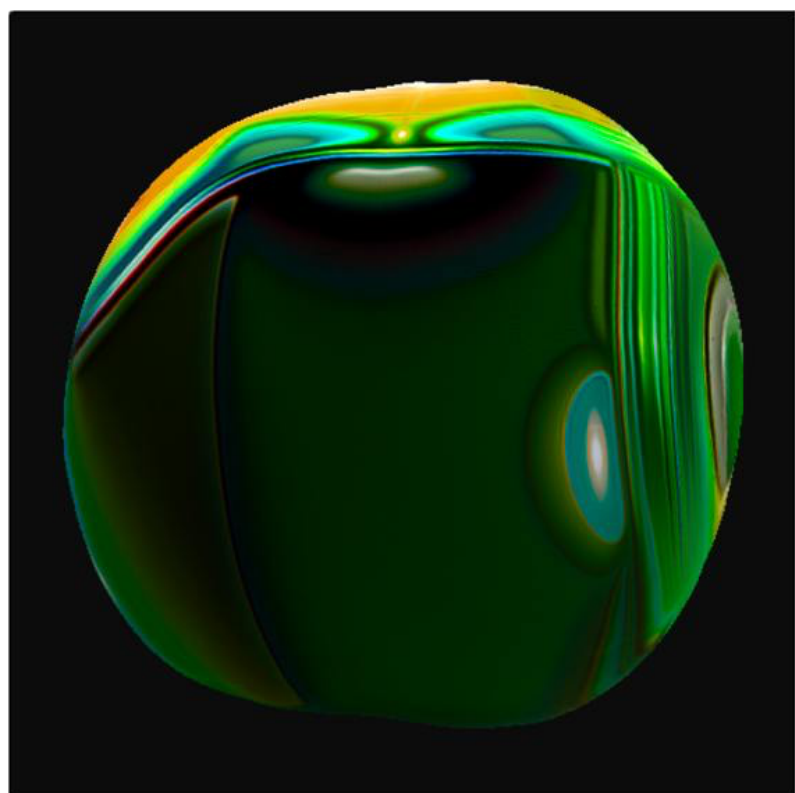
ISM

HiFA

Ours

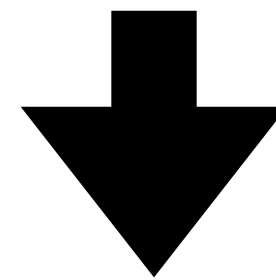


Опыт 3: о пользе `torch.clamp()`



Перепараметризация SDS

$$\nabla_{\theta} \mathcal{L}_{\text{SDS}}(\theta) \approx \mathbb{E}_{t, \epsilon, c} \left[\underbrace{\omega(t) (\epsilon_{\phi}(\mathbf{x}_t, t, y) - \epsilon)}_{\text{SDS update direction}} \frac{\partial g(\theta, c)}{\partial \theta} \right]$$

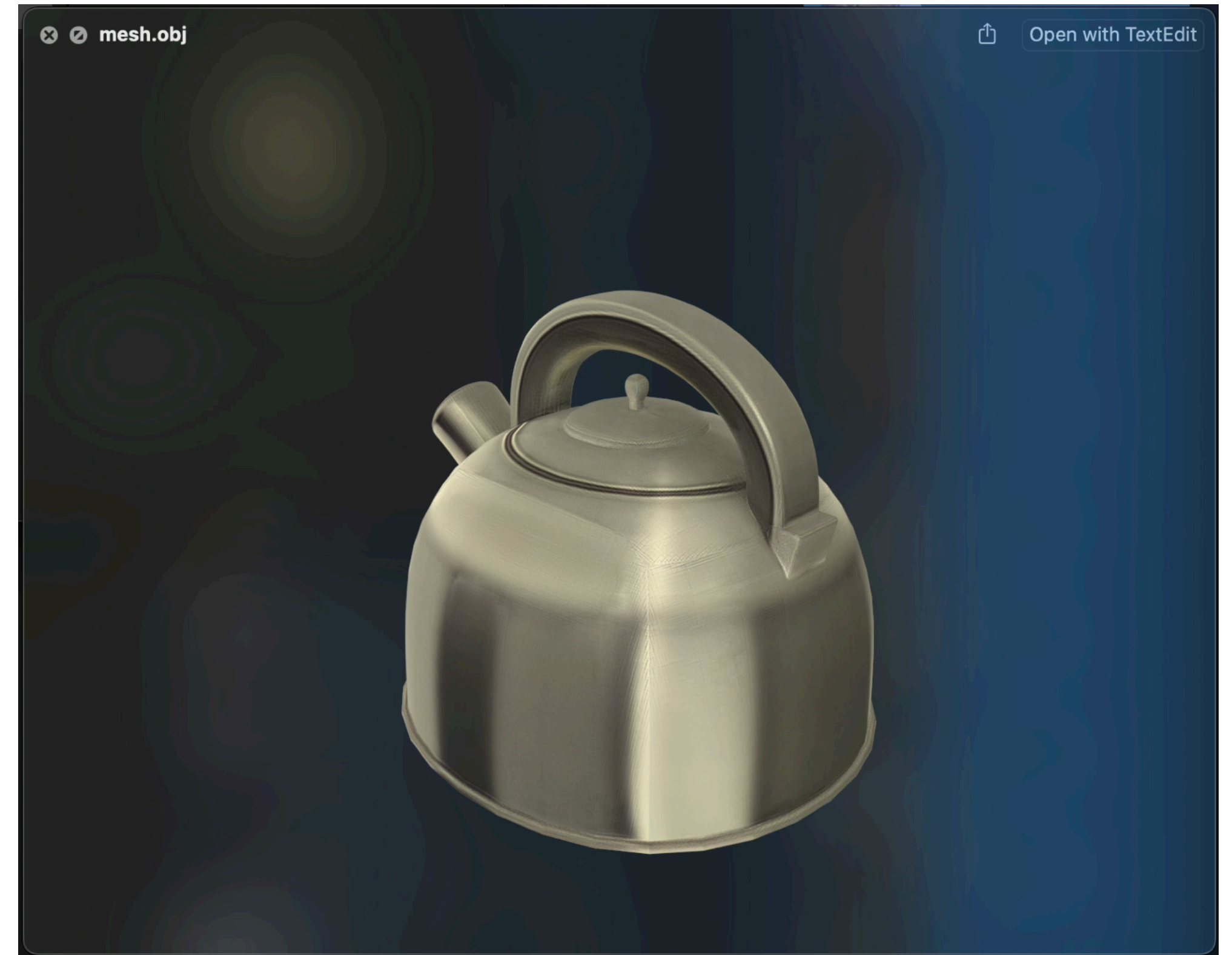
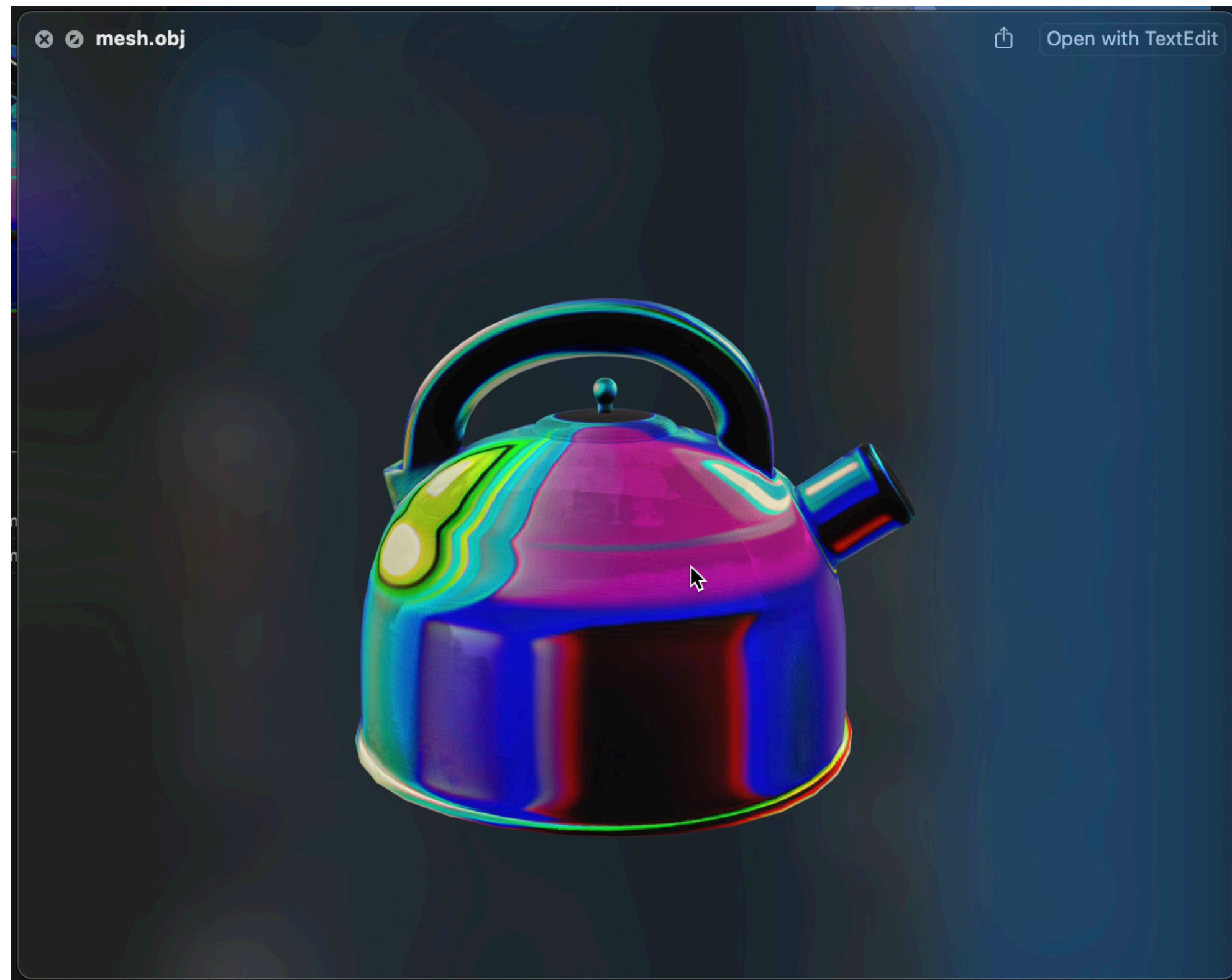


$$\nabla_{\theta} \mathcal{L}_{\text{SDS}}(\theta) = \mathbb{E}_{t, \epsilon, c} \left[\frac{\omega(t)}{\gamma(t)} (\mathbf{x}_0 - \hat{\mathbf{x}}_0^t) \frac{\partial g(\theta, c)}{\partial \theta} \right]$$

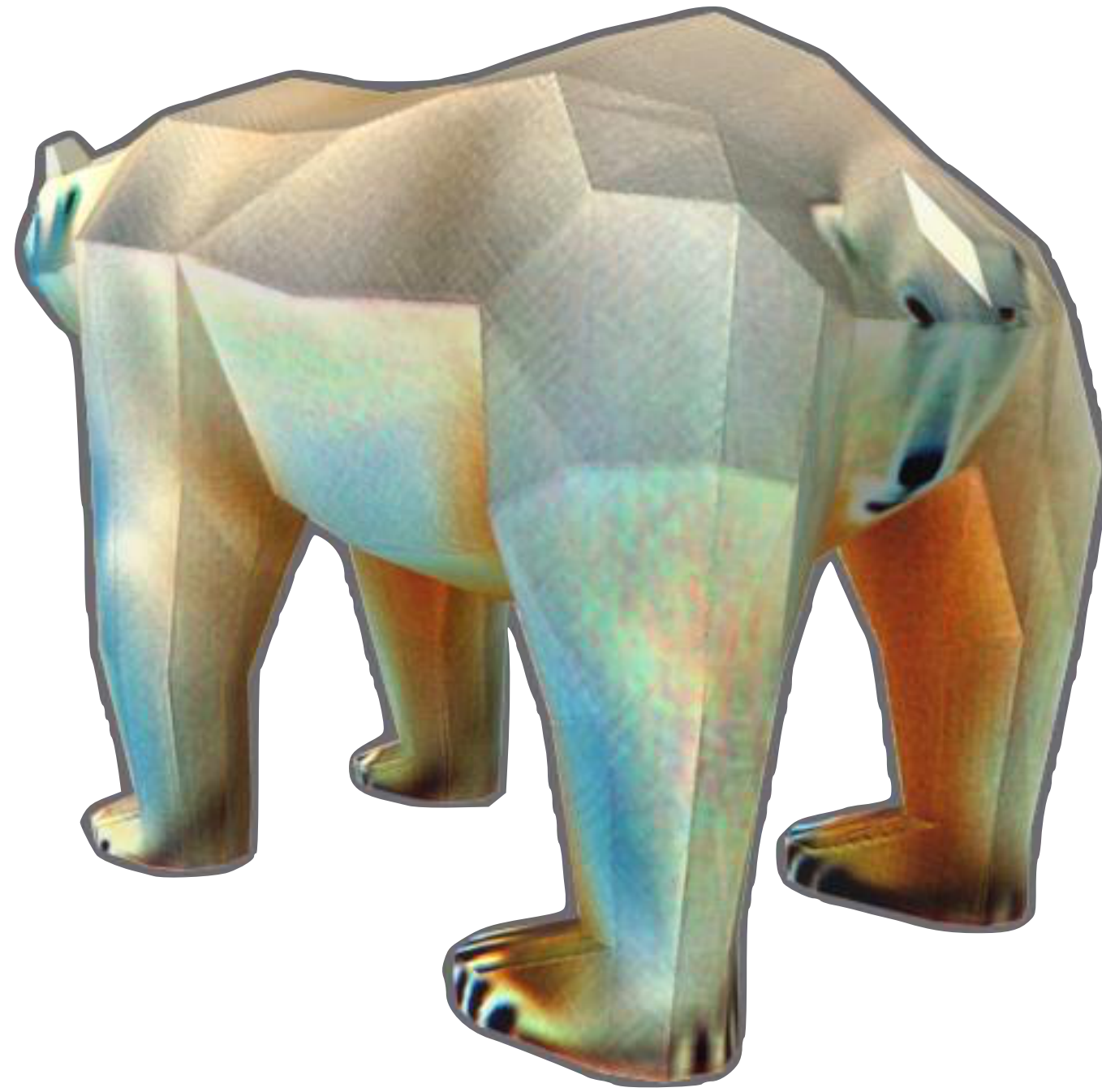


Clamp it!

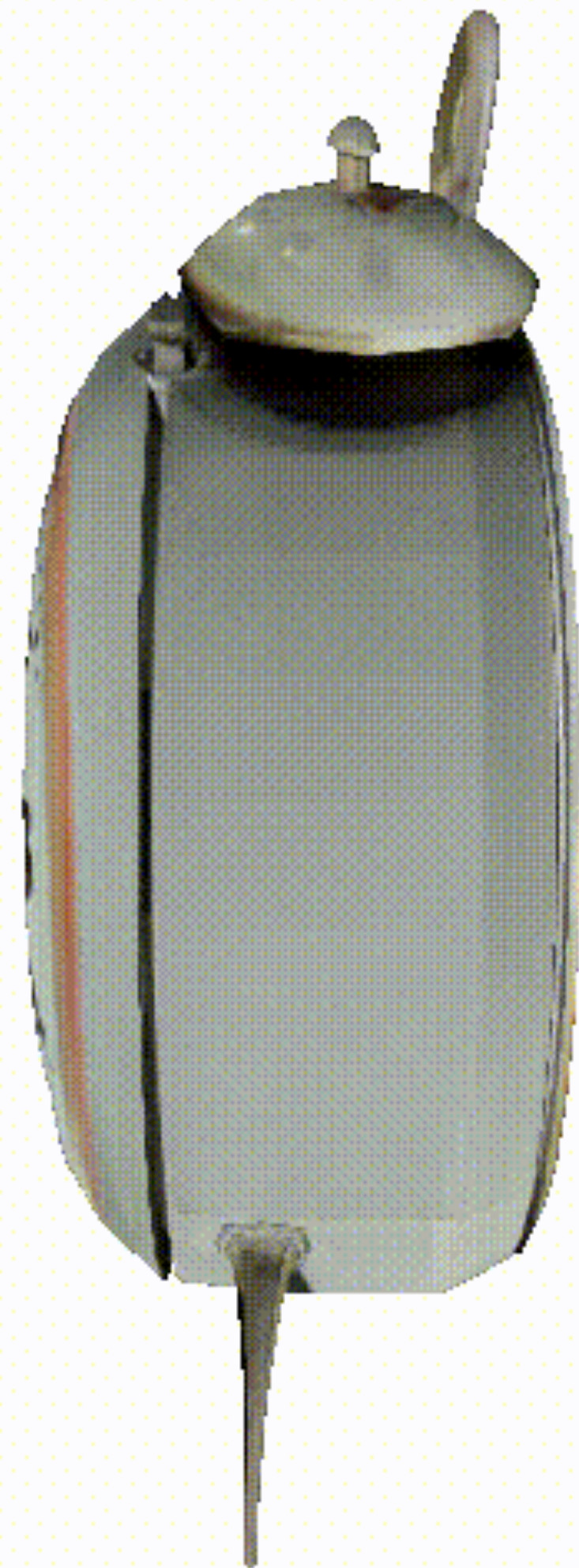
Перепараметризация SDS



Еще картинки







Мысли про будущее

- Аккуратнее изучить влияние освещения на результаты
- Уходить от SDS
 - Неестественные цвета
 - Ньюансы освещения не удастся достаточно надежно передать
 - С латентными моделями в нашей задаче он не дружит
- Более аккуратный контроль сигнала от диффузионок
 - Дообучение с учётом ракурса, карт глубин и освещения