Text2Tex

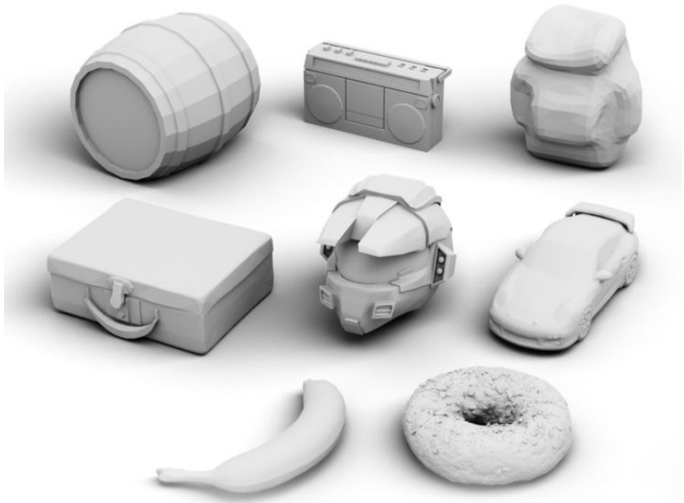
Text-driven Texture Synthesis via Diffusion Models

Писцов Георгий 23.10.23

О чем сегодня поговорим

- Постановка задачи
- Описание метода
- Depth-aware image inpainting
- Dynamic view partitioning
- Refinement/Automatic viewpoint selection
- Результаты
- User/Ablation study
- Артефакты

Постановка задачи



Meshes without textures

Generated textures with text prompts

Описание метода

- Авторы используют **Denoising Diffusion Probabilistic Model**
- Сначала кодируем в латентное пространство и получаем z_0. Дальше пропускаем через марковскую цепь

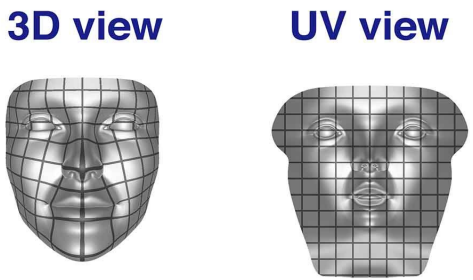
$$z_t \sim \mathcal{N}(\sqrt{1-\beta_t} z_{t-1}, \beta_t \mathbf{I}). \qquad z_t \sim \mathcal{N}(\sqrt{\bar{a}_t} z_0, (1-\bar{a}_t)\mathbf{I}), \qquad \bar{a}_t = \sum_{t=i}^{T} (1-\beta_t)$$

$$\hat{z}_{t-1} \sim \mathcal{N}(\mu_\theta(z_t, t), \sigma_\theta(z_t, t))$$

- Чтобы не делать полный рандом они вводят гамму (**denoising strength**) от 0 до 1. Параметр контролирует количество шагов диффузии, по сути мы на него просто умножаем

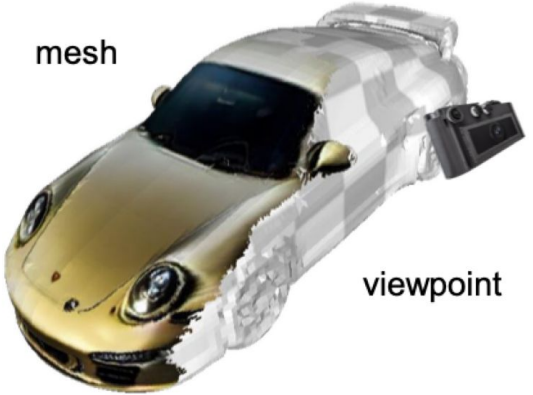
Depth-aware image inpainting

- Главная цель генерации текстур это закрашивать пропущенные регионы
- Авторы используют предобученную Depth2Image модель
- Чтобы не генерировать изображение целиком, они встраивают inpainting mask в процесс диффузии
- Чтобы превратить изображение(взгляд из точки) в текстуру авторы используют UV параметризацию



Dynamic view partitioning: Similarity mask

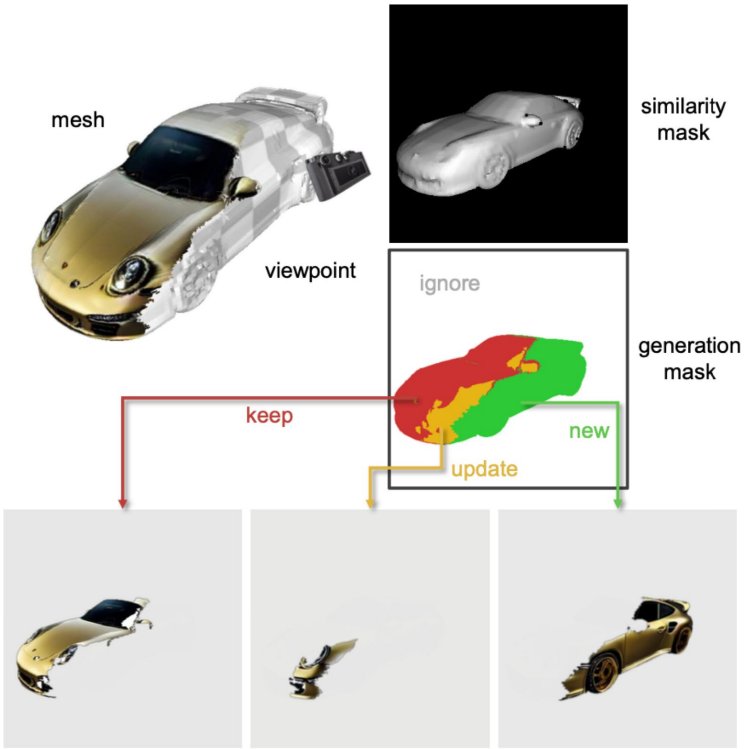
- Каждый пиксель в similarity mask представляет собой обратное нормированное значение cosine similarity между векторами нормалей видимых поверхностей и направлением взгляда.
- В целом эти маски показывают степень поворота лица от точки обзора



similarity mask

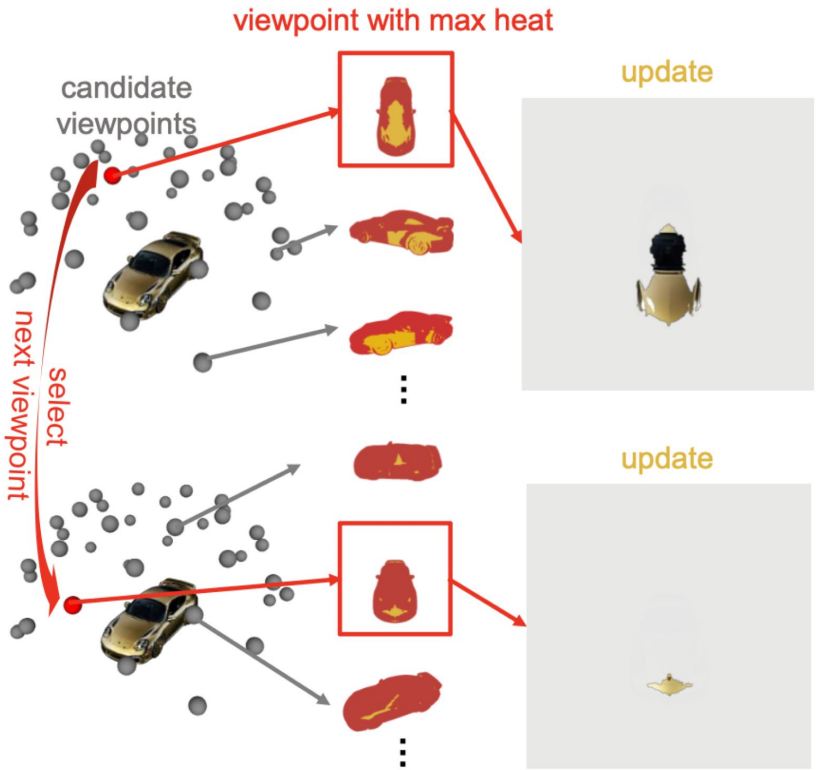
Dynamic view partitioning : Generation mask

- **New**: не нанесена текстура
- **Update**: score в similarity mask выше, чем во всех остальных ракурсах
- **Keep**: не наивысший score в similarity mask среди ракурсов
- **Ignore**: фон

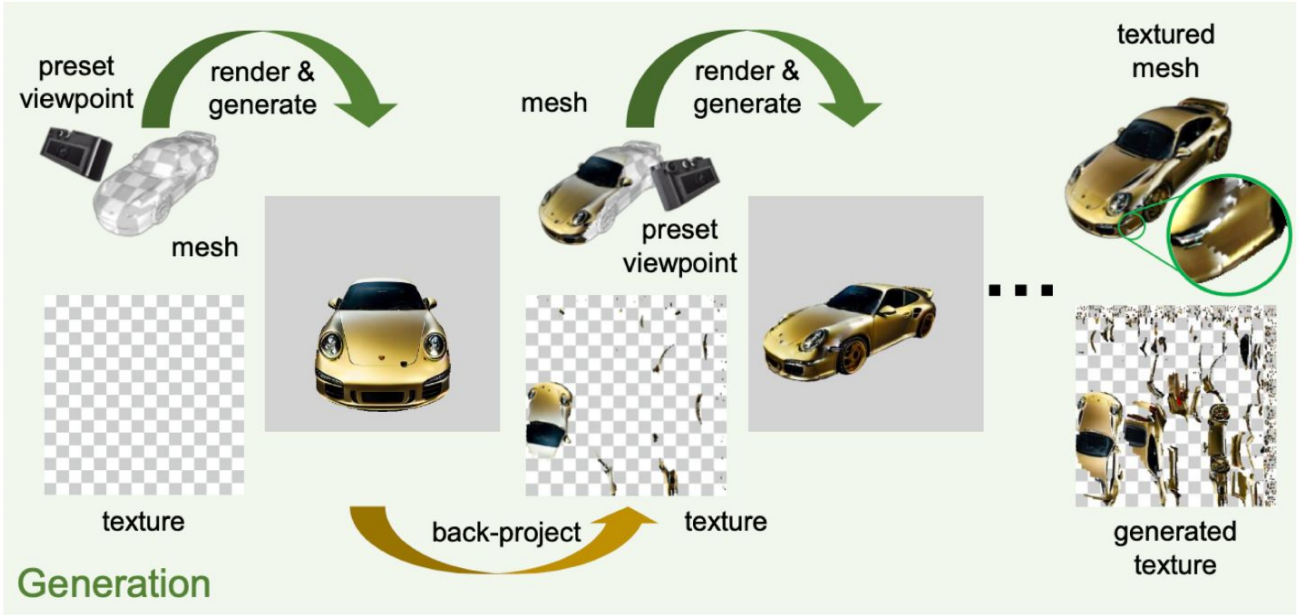


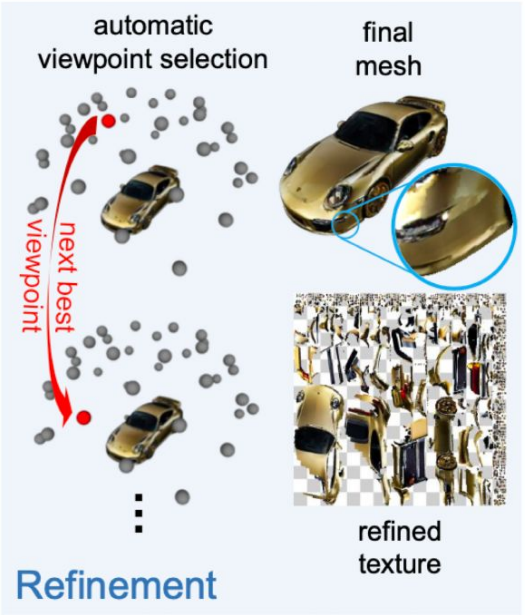
Refinement: automatic viewpoint selection

- Считаем heat - представляет собой нормированную площадь update относительно текущей видимой области объекта.
- Выбираем следующую точку по argmax heat
- Обновляем текстуры с небольшим denoising strength



Общий пайплайн решения





Эксперименты

Датасет:

- Они берут подмножество датасета Objaverse, семплируя по 3 объекта из каждой категории
- Авторы вручную убирают нерепрезентативные объекты
- После всех манипуляций у них остается 410 объектов суммарно в 225 категориях
- Для сравнения с GAN они берут 300 объектов “car” из ShapeNet

Метрики:

- Frechet Inception Distance (FID)
- Kernel Inception Distance (KID)

С чем сравнивали?

Quantitative results

Method	FID ↓	KID (x10^{-3}) ↓
Text2Mesh [34]	45.38 (+9.7)	10.40 (+2.7)
CLIPMesh [37]	43.25 (+7.6)	12.52 (+4.8)
Latent-Paint [33]	43.87 (+8.1)	11.43 (+3.7)
Text2Tex (Ours)	**35.68**	**7.74**

Table 1: Quantitative comparisons on Objaverse subset. Our method performs favaorably against state-of-the-art text-driven texture synthesis methods.

Method	FID ↓	KID (x10^{-3}) ↓
Texture Fields [42]	177.15 (+130.2)	17.14 (+12.8)
SPSG [16]	110.65 (+63.7)	9.59 (+5.2)
LTG [65]	70.76 (+23.8)	5.72 (+1.4)
Texturify [56]	59.55 (+12.6)	4.97 (+0.6)
Text2Tex (Ours)	**46.91**	**4.35**

Table 2: Quantitative comparison on the ShapeNet cars. Our method outperform state-of-the-art category-specific GAN-based methods by a significant margin.

Figure 6: Qualitative comparisons on ShapeNet car. Our method generates sharper and more coherent textures with respect to the geometries compared to the state-of-the-art GAN-based method.

User study: ставим эксперименты на людях

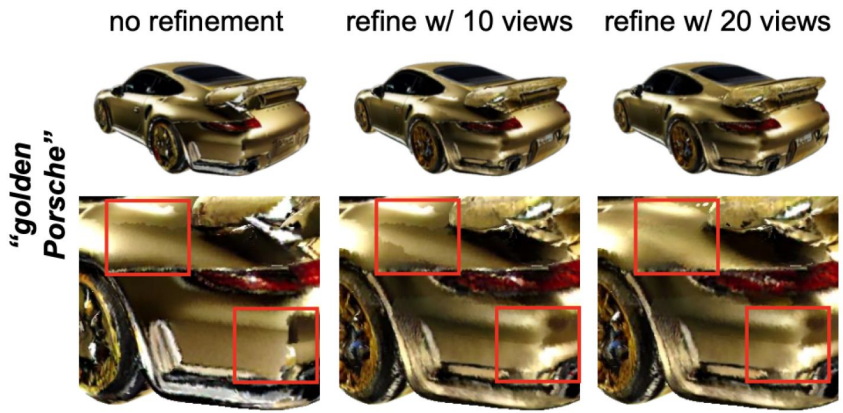
- Авторы проводят side-by-side тестирования для каждого из бейзланов и своего метода.
- В итоге они собрали 604 ответа от 41 пользователя.
- По сравнению с CLIPMesh и Text2Mesh, метод авторов предпочитают 83,92% и 76,47% соответственно.
- 64,18% склоняются к методу автора, чем к конкурирующему Latent-Paint.

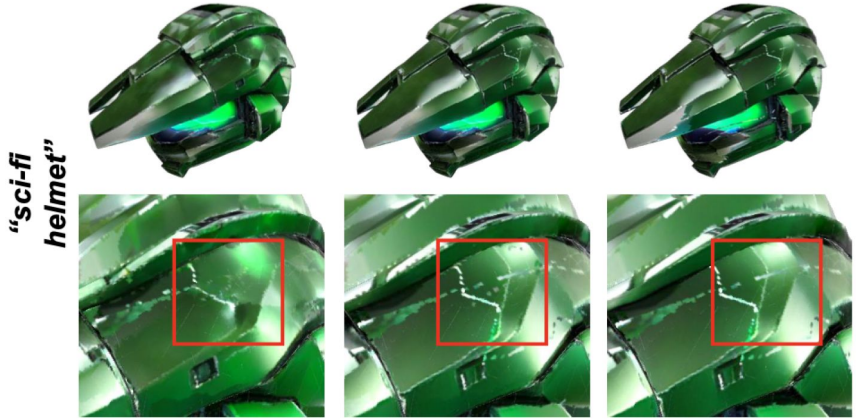
Ablation study: depth-aware inpainting & update

w/ Depth2Img	w/ inpainting	w/ update	FID ↓	KID ($\times 10^{-3}$) ↓
✓	x	x	39.88	9.78
✓	✓	x	38.19	9.11
✓	✓	✓	**37.09**	**8.78**

Ablation study: refinement

# views	0	5	10	15	20
↓ FID	37.09	36.67	36.39	35.98	**35.68**
↓ KID (x10^{-3})	8.78	8.31	8.12	7.98	**7.74**





Артефакты

- Несмотря на способность создавать высококачественные 3D-текстуры, метод склонен создавать текстуры с эффектами затенения.
- Проблема может быть решена путем тщательной и точной настройки входных запросов, однако это требует дополнительных усилий со стороны человека и не может быть хорошо масштабировано на массивные объекты генерации.
- Одним из возможных решений является точная настройка диффузионной модели для устранения затемнения текстур.

Приколы

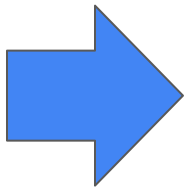
Figure 12: Different styles for the Porsche. Our method is capable of handling complicated styles such as “baroque” and “cyberpunk” without distorting the original properties of the input geometry.

Приколы

Figure 13: Creative textures for the Porsche with unrealistic prompts. Our method clearly represents the original properties of the geometry, while reflecting iconic characteristics of the input prompts.

Чего не хватило мне?

“Sally Carrera”



Вопросы?