Применение генеративных моделей для аугментации обучающих данных в задачах сегментации изображений

Изместьева У.А.
izmesteva.ua@phystech.edu

2 курс магистратуры «Методы и технологии искусственного интеллекта» Центр когнитивного моделирования МФТИ

2024 – Научно-исследовательский семинар

Мотивация

Основания разработки:

- **Расширение данных.** Наборы разнообразных данных повышают устойчивость моделей при обучении.
- **Узкоспециализированные сценарии.** Моделирование специфичных ситуаций помогает решать задачи с ограничениями данных.
- **Снижение стоимости.** Ручной сбор и разметка данных дорогостоящая и подвержена человеческим ошибкам.

Цель: разработать инструмент для автоматической аугментации обучающих данных на основе генеративной модели, способной создавать синтетические данные в заданном домене используя реальные данные и различные модальности.

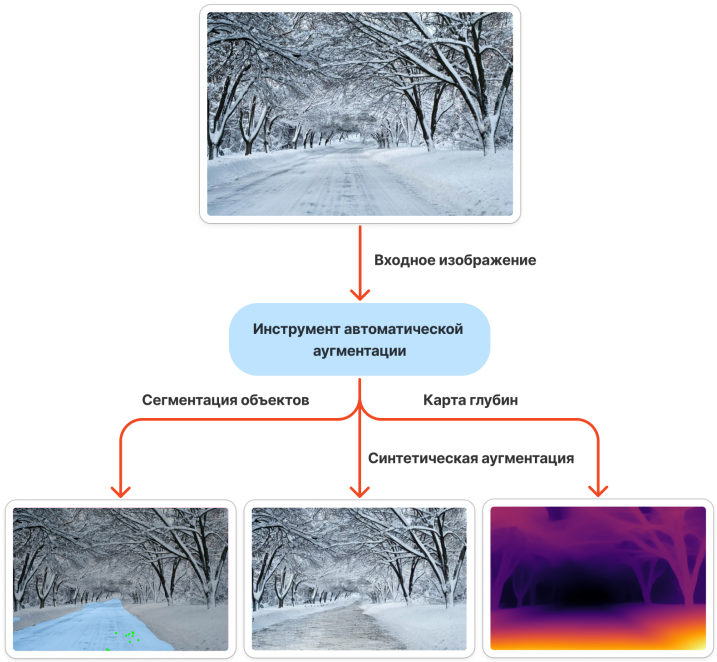
Постановка задачи

Входные данные:

- набор размеченных изображений;
- карты глубин (опционально);
- текстовые описания объектов (опционально).

Выходные данные:

- аугментированные изображения;
- аугментированные карты глубин;
- список масок, представляющих объекты на входных изображениях (опционально);



Постановка задачи

Задача состоит в разработке модели F, которая по входным данным X генерирует выходные данные Y. F: $X \rightarrow Y$, где:

- X содержит изображения *(I)*, маски объектов *(M)*, текстовые классы объектов *(T)*;
- Y представляет измененные изображения*(X')* и карты глубины *(D')*.

Математическая формулировка

Пусть $F_\theta(\hat{c})$ - это модель, которая отображает условие текста и маски объекта $\hat{c}$ в условный вектор. Тогда, представление функции потерь $L_{inpaint}$ имеет вид:

$$L_{inpaint} := \mathbb{E}_{z\sim E(x),\hat{y},\epsilon\sim \mathrm{N}(0,1),t}[\|\epsilon - \epsilon_\theta(z_t, t, F_\theta(\hat{c}))\|^2] \tag{2}$$

где:

- t - шаг во времени;
- z_t - латентное изображение, зашумленное в момент времени t;
- ϵ - образец немасштабированного шума;
- θ - нейронная сеть расшумления.

Научная новизна

Научная новизна работы:

- **Функция потерь:** штрафуем модель при попытке сгенерировать "пустоту"и поощряем за генерацию объектов

$$L_{new} := \alpha \|\epsilon_{obj} - \epsilon_\theta(z_t, t, F_\theta(\hat{c}))\|^2 + \frac{\beta}{\|\epsilon_{empt} - \epsilon_\theta(z_t, t, F_\theta(\hat{c}))\|^2 + \gamma} \quad (3)$$

где:

  - ϵ_{obj} - шум GT объекта;
  - ϵ_{empt} - шум синтетически сгенерированной пустоты;
  - $F_\theta(\hat{c})$ - выход модели;
  - α и β - весовые коэффициенты;
  - γ - обработка исключения деления на ноль.

- **Пайплайн генерации синтетических датасетов:** на основе синтетической карты глубины генерировать сцены в желаемом домене + генерация специфических объектов;

Пайплайн генерации синтетических датасетов



Общая схема

Постановка экспериментов

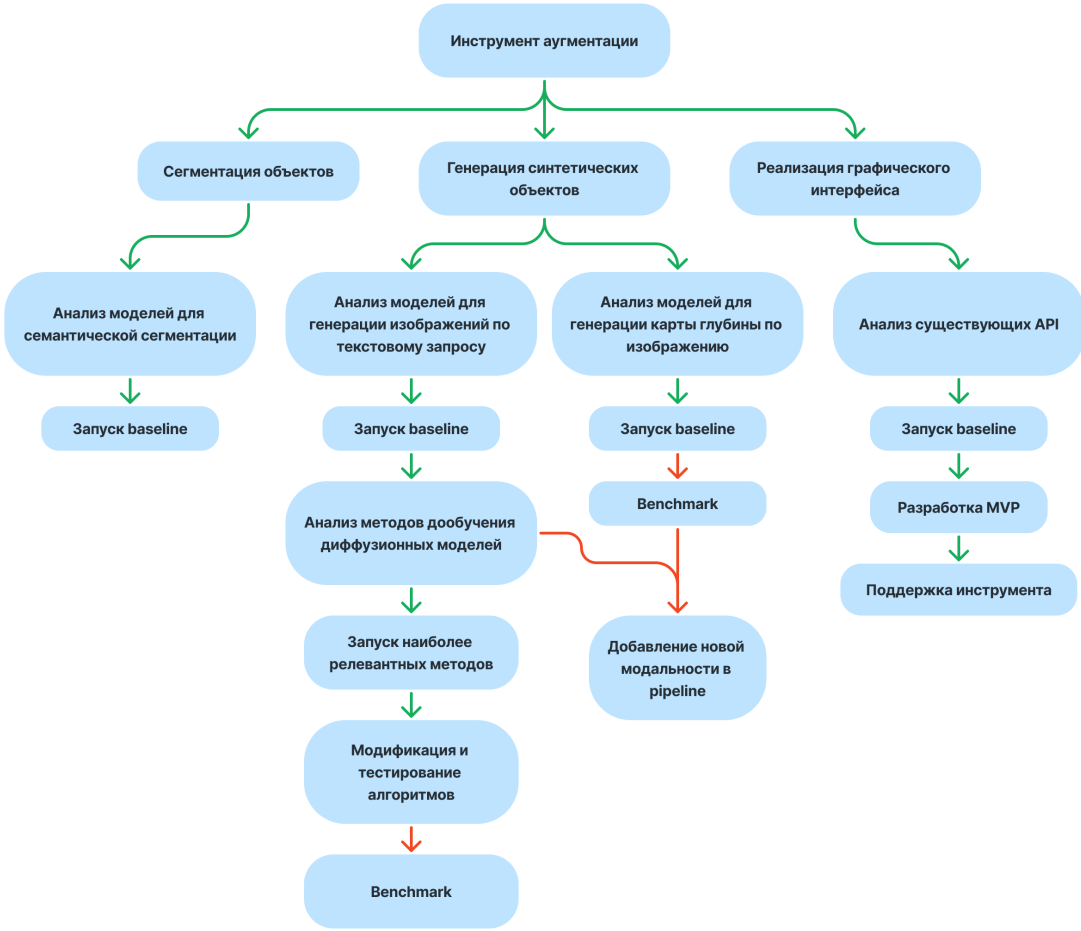


Схема проведенных экспериментов

Стек технологий. Генерация объектов

Генерация изображений в режиме inpainting:

- Stable Diffusion 2[1]
- PowerPaint[2]

Дообучение диффузионных моделей:

- Textual inversion[3]
- DreamBooth[4]
- LoRA[5]

[1] Rombach R. et al. High-resolution image synthesis with latent diffusion models. – 2022.

[2] Zhuang J. et al. A task is worth one word: Learning with task prompts for high-quality versatile image inpainting – 2023.

[3] Gal R. et al. An image is worth one word: Personalizing text-to-image generation using textual inversion – 2022.

[4] Ruiz N. et al. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation – 2023.

[5] Hu E. J. et al. Lora: Low-rank adaptation of large language models – 2021.

Стек технологий. Вычисление глубины

- Depth Anything[6]
- Depth Anything V2[7]
- MaryGold[8]
- DPT[9]

[6] Yang L. et al. Depth anything: Unleashing the power of large-scale unlabeled data – 2024.

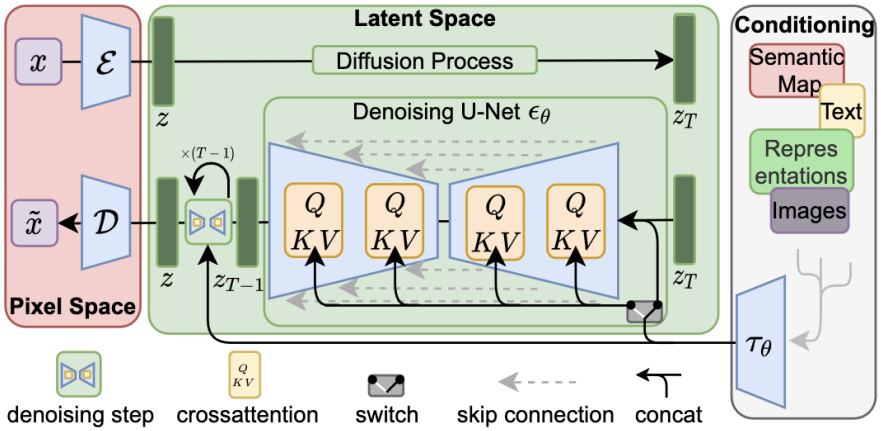
[7] Yang L. et al. Depth Anything V2 – 2024.

[8] Ke B. et al. Repurposing diffusion-based image generators for monocular depth estimation – 2024.

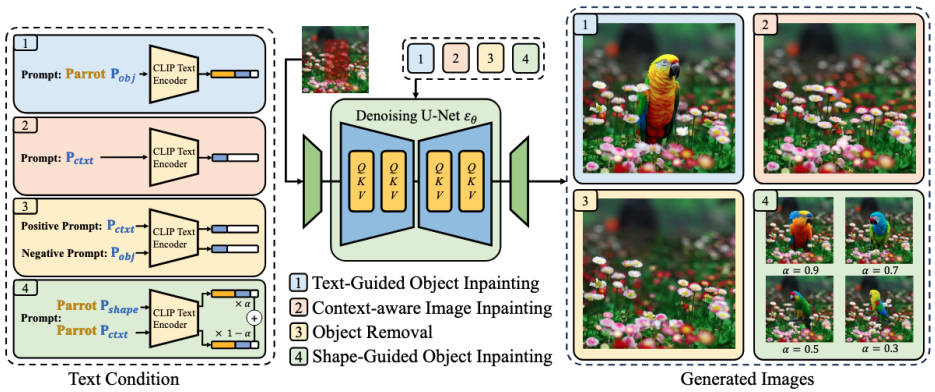
[9] Ranftl R., Bochkovskiy A., Koltun V. Vision transformers for dense prediction – 2021.

Используемые методы. Диффузионные модели

Генерация изображений в режиме inpainting - процесс восстановления или восполнения участков изображения по выделенной маске.



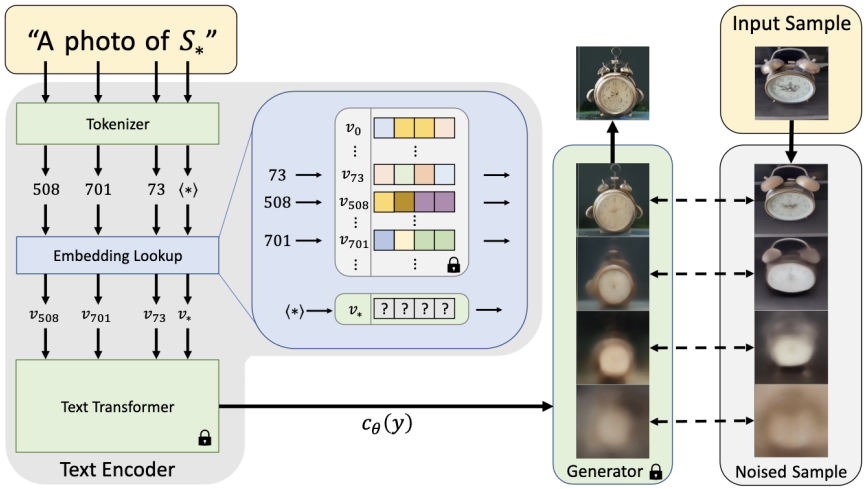
Stable Diffusion 2



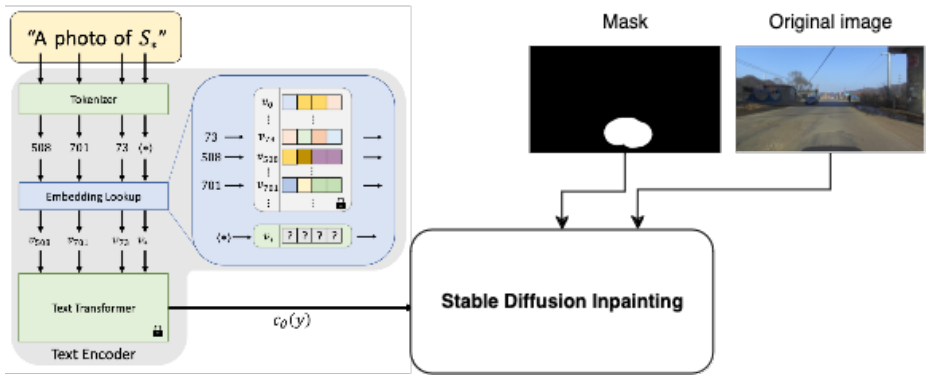
PowerPaint

Используемые методы. Textual inversion

Идея: использовать inpainting веса для генерации объектов в заданной области



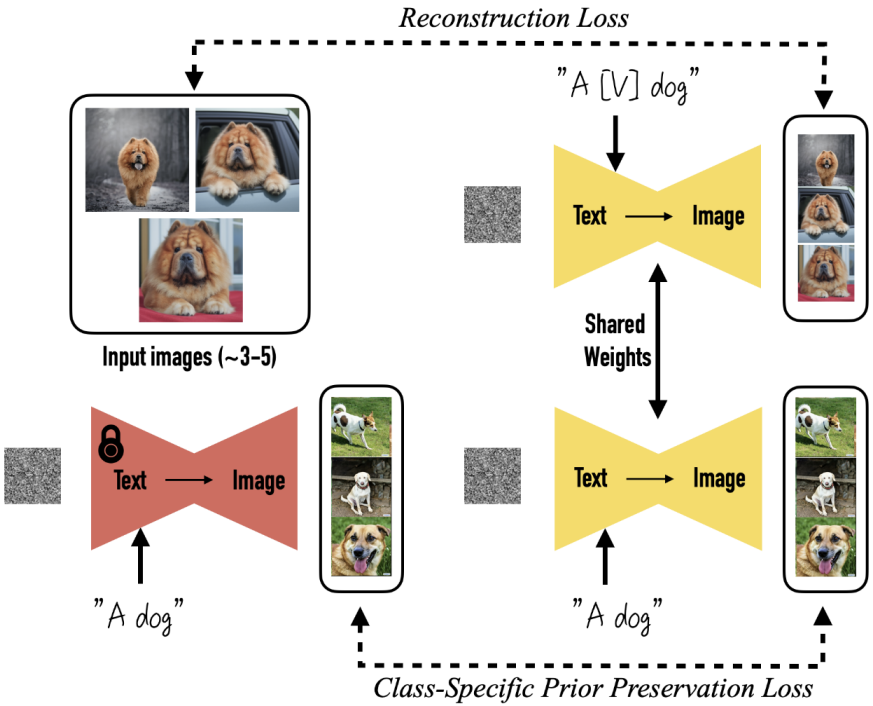
Текстовая инверсия



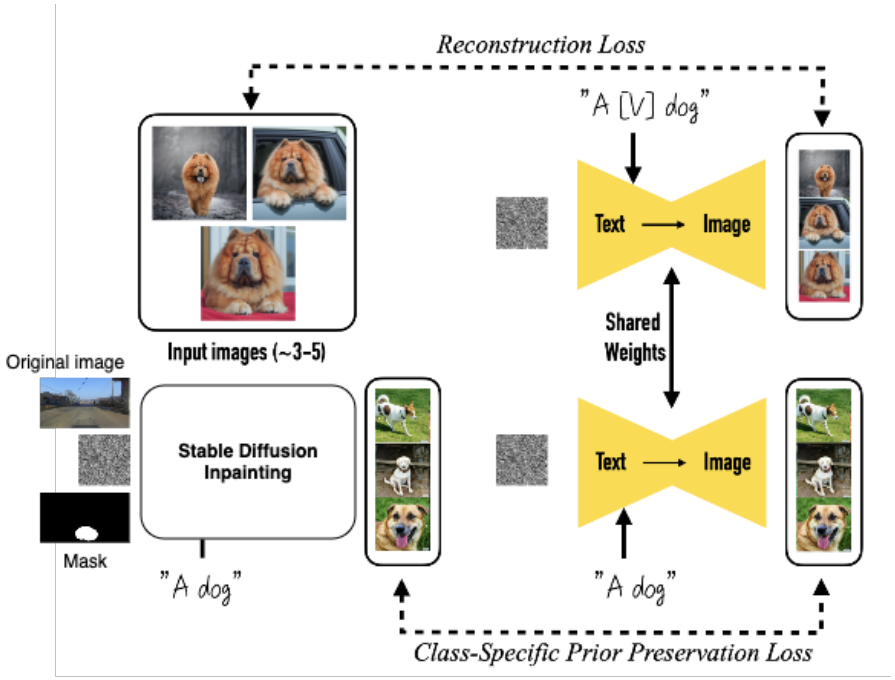
Модифицированный алгоритм

Используемые методы. DreamBooth

Идея: использовать inpainting веса для генерации объектов в заданной области

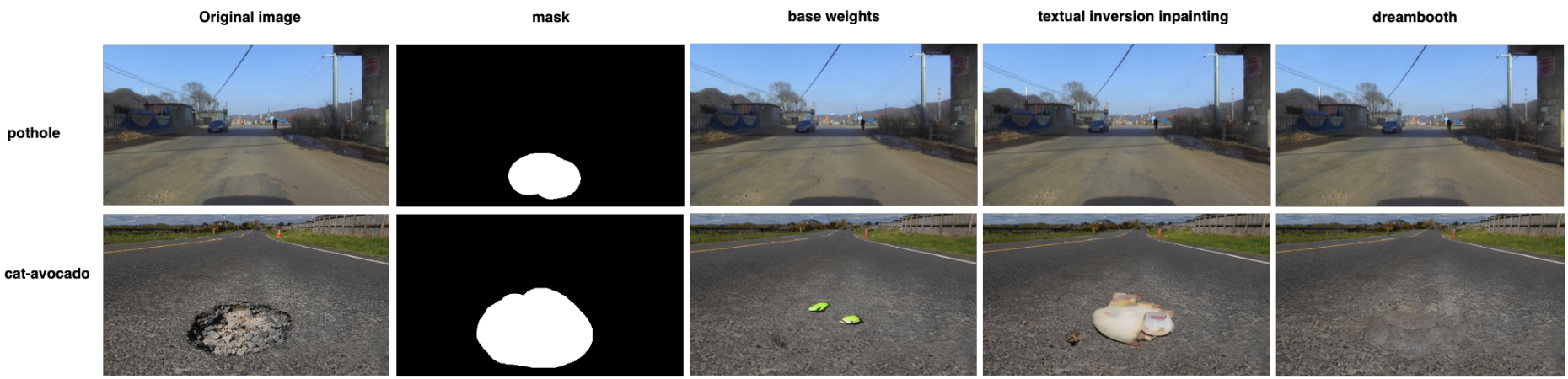


DreamBooth



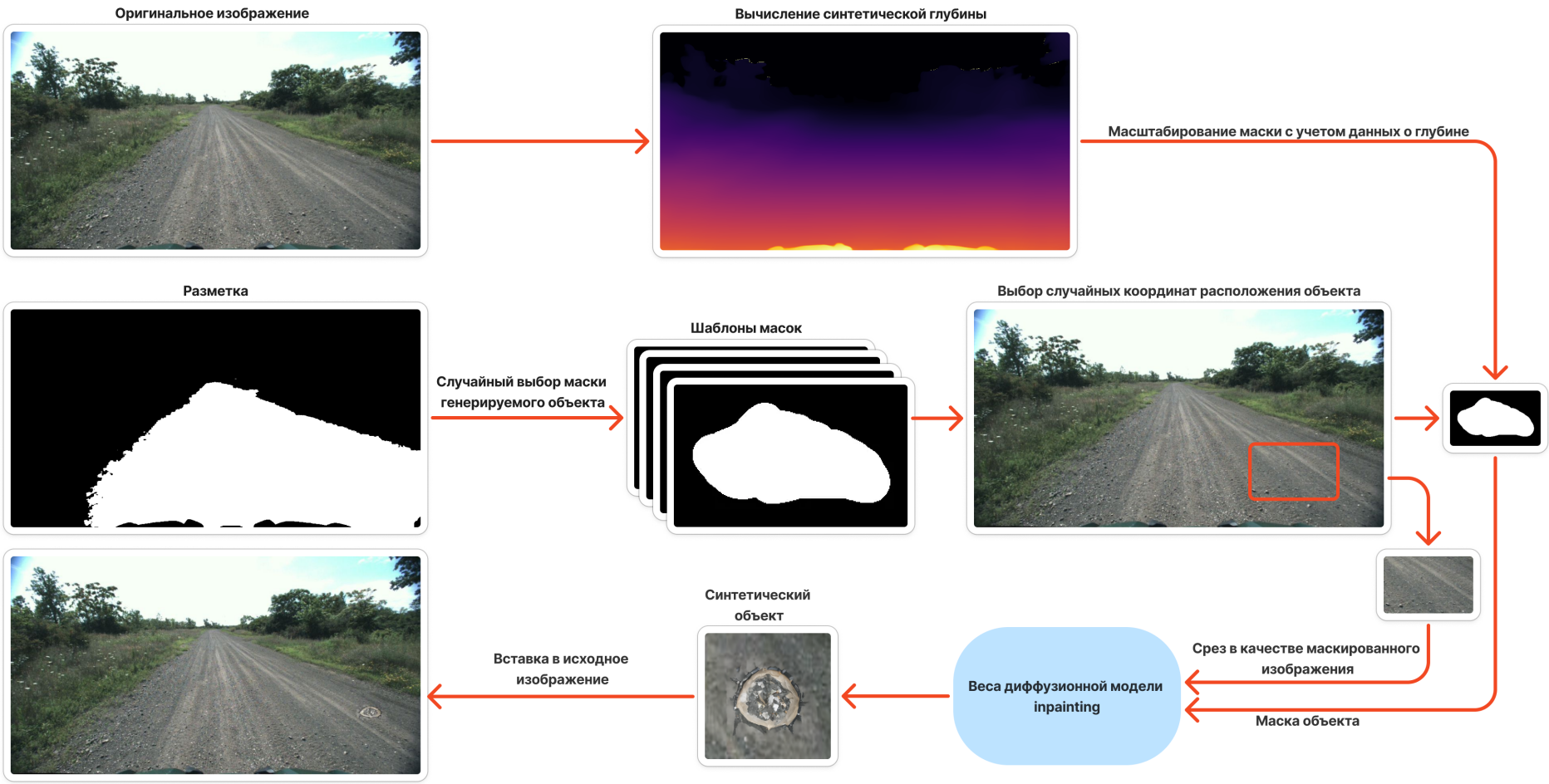
Модифицированный алгоритм

Эксперименты. Генерация объектов

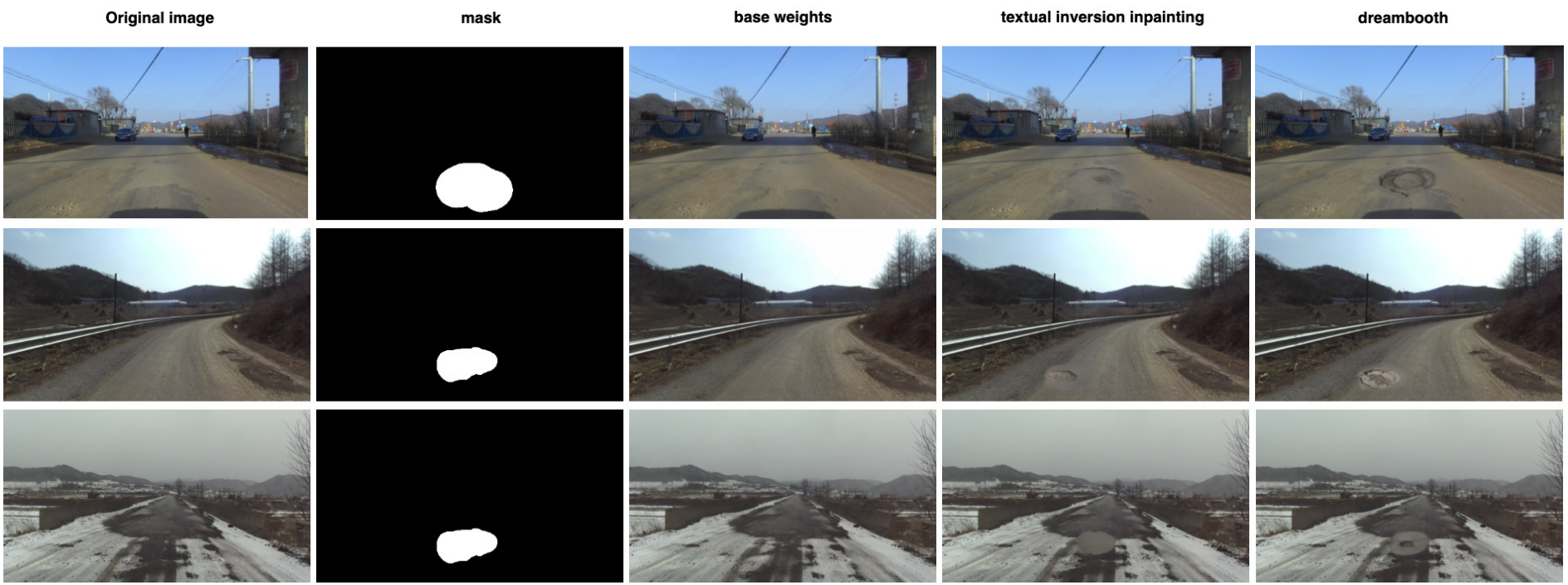


Результаты экспериментов с использованием весов Stable Diffusion 2

Пайплайн генерации синтетических объектов



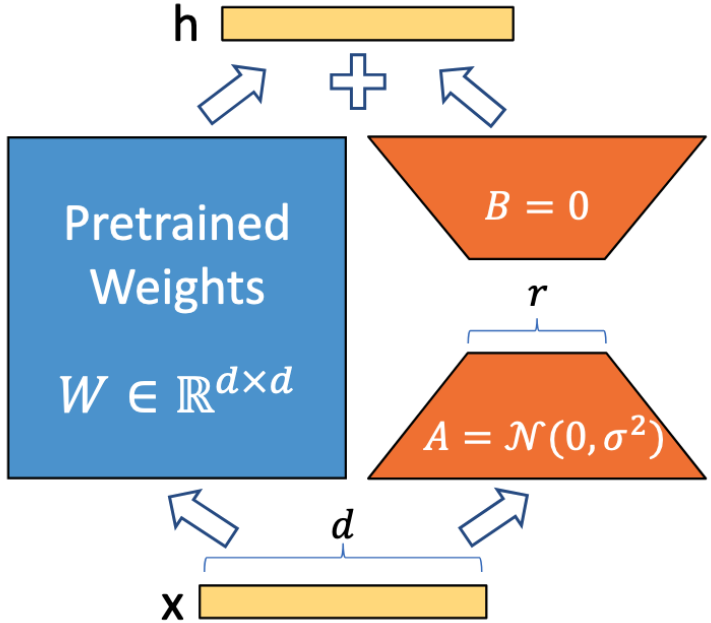
Эксперименты. Генерация объектов



Результаты экспериментов с использованием весов Stable Diffusion 2 и кропов изображений, включающих маску объекта

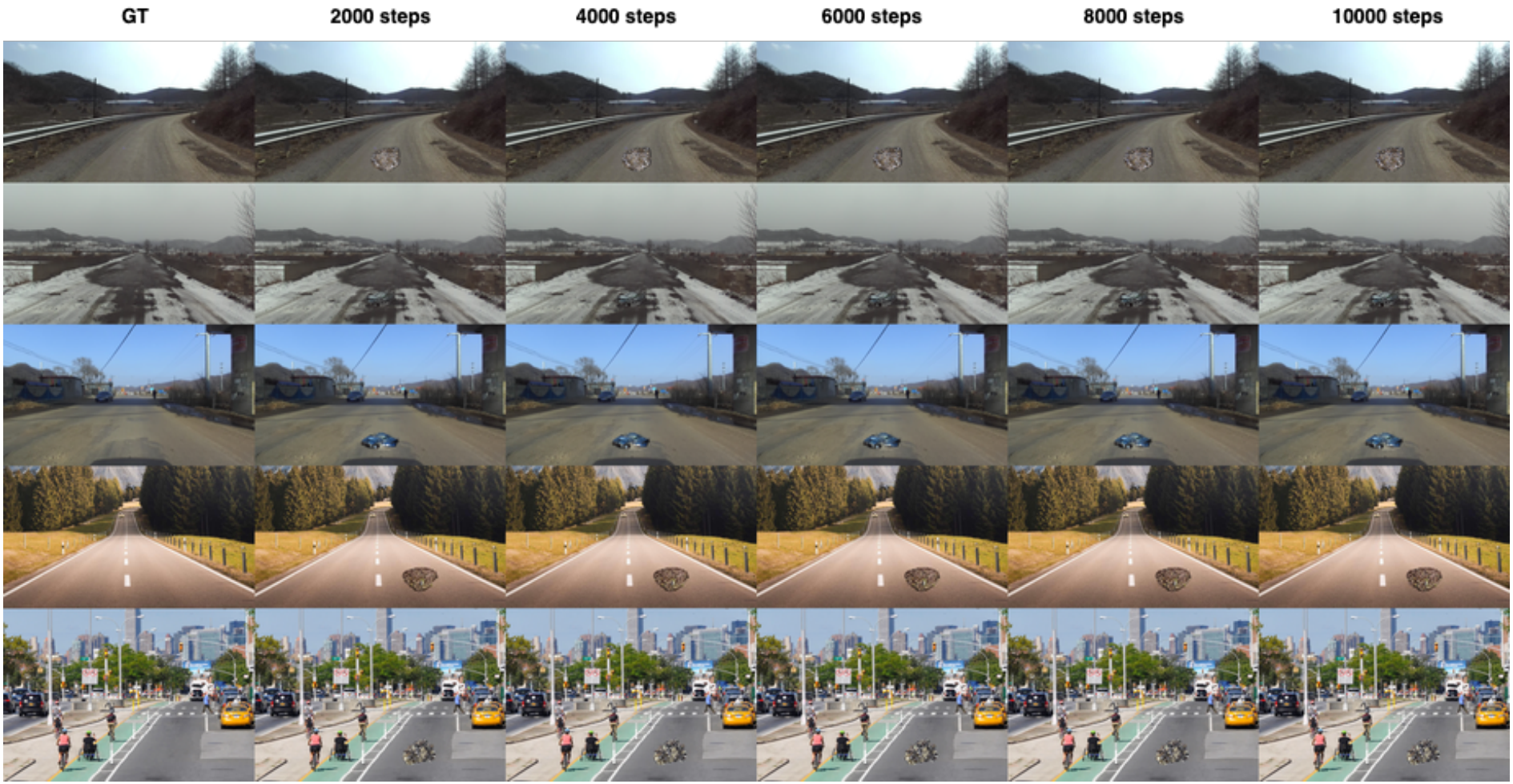
Используемые методы. LoRA

Идея: использовать inpainting веса для генерации объектов в заданной области



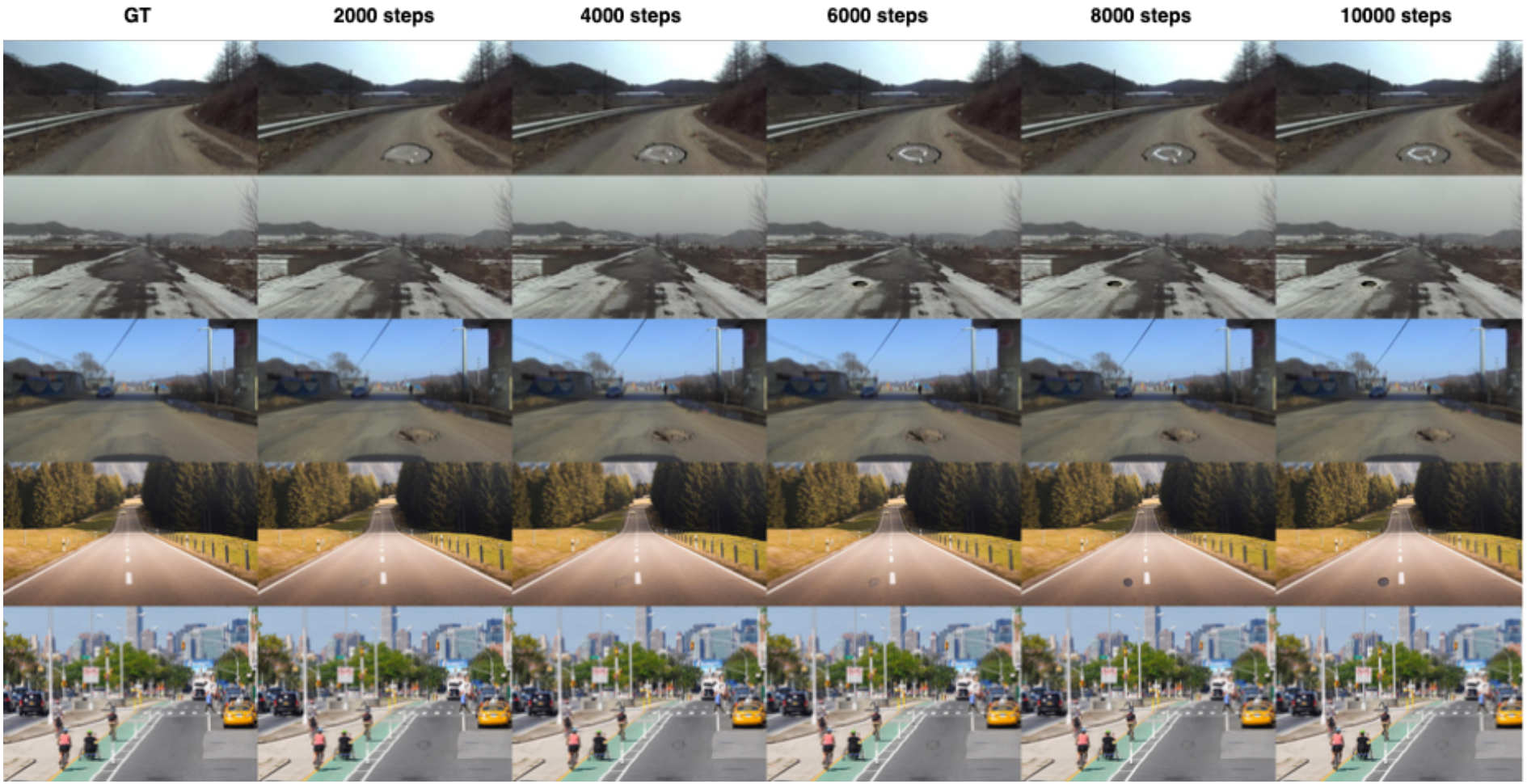
LoRA

Эксперименты. Генерация объектов



Результаты экспериментов с использованием весов PowerPaint и кропов изображений, включающих маски объекта

Эксперименты. Генерация объектов



Результаты экспериментов с использованием весов Stable Diffusion 2 и кропов изображений, включающих маски объекта

Сравнение моделей глубины

Название модели	**Среднее время (с)**	**RMSE**	**AbsRel**
Depth Anything	0.0380	9.1094	7.1462
Depth Anything V2	0.0384	8.9629	7.1368
MariGold	0.4283	10.4906	9.7801
DPT	0.0689	10.3837	6.9439

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(z_i - \hat{z}_i)^2}, \quad \text{AbsRel} = \frac{1}{N}\sum_{i=1}^{N}\frac{|z_i - \hat{z}_i|}{z_i}$$

где z_i — истинное значение глубины, $\hat{z}_i$ — предсказанное значение.

Эксперименты. Использование глубины



Результаты экспериментов с использованием весов Depth Anything V2 на датасете INSTSnowRoadDetection

Выводы

- Лучший метод тюнинга: Stable Diffusion 2 + LoRA
- Наиболее подходящий метод вычисления глубины - Depth Anything V2

Заключение

Были осуществлены следующие этапы:

- Сбор и разметка датасета со специфичными объектами
- Реализация и тестирование специализированного алгоритма LoRA для задачи inpainting
- Бенчмаркинг моделей вычисления глубины

Дальнейшие исследования:

- Реализация обучения с использованием новой функции потерь
- Генерация положительных препятствий
- Совмещение пайплайнов генерации новых объектов и создания сцен на основе карты глубины

Спасибо за внимание!

izmesteva.ua@phystech.edu