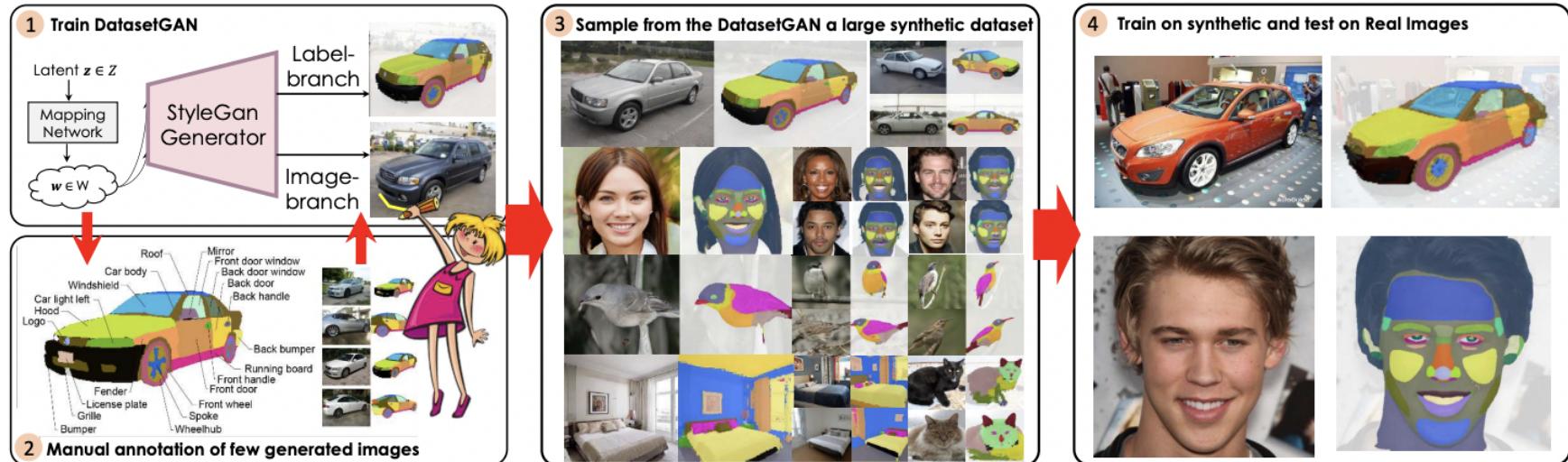


# DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort

Илюхов Алексей  
Гришанин Виктор  
Исаев Сергей  
Виноградова Дарья  
13 октября 2021

# DatasetGAN

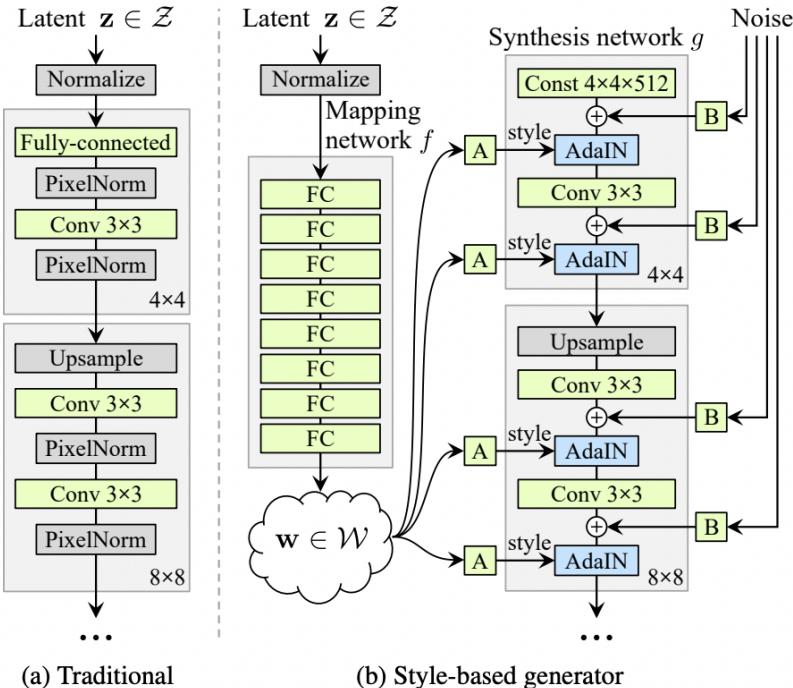


# Архитектура DatasetGAN



В качестве генератора используем StyleGAN, в качестве интерпретатора архитектуру из статьи

# Основные идеи StyleGAN

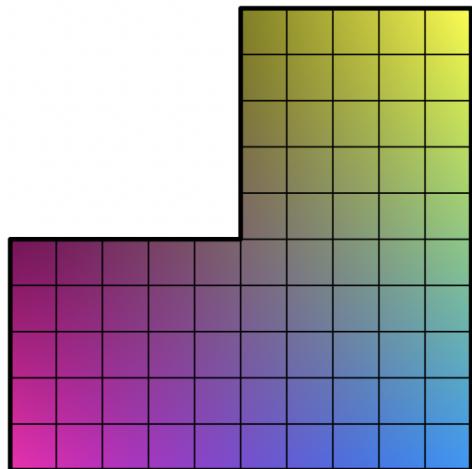


Используем  $z$  не в качестве стартового состояния, а для генерации стиля

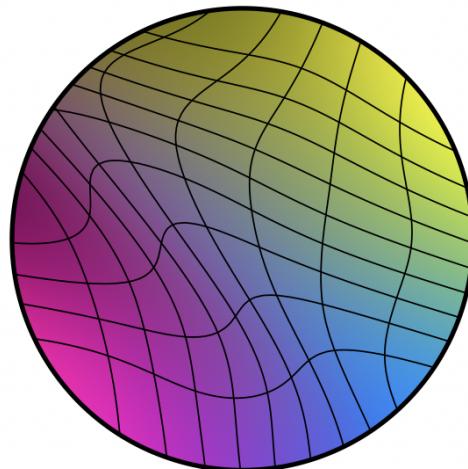
Преобразуем  $z$  полносвязной сетью

Для применения стиля используем AdaIN

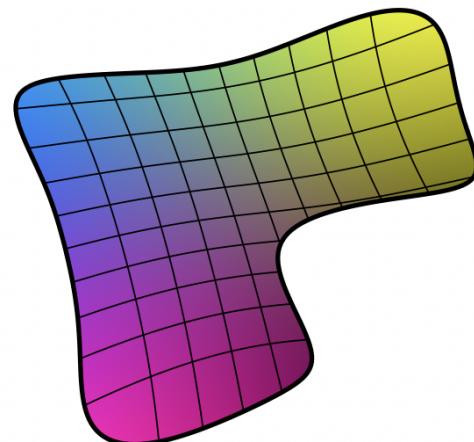
Зачем нужна полносвязная сеть в начале



(a) Distribution of features in training set



(b) Mapping from  $\mathcal{Z}$  to features



(c) Mapping from  $\mathcal{W}$  to features

## Добавление стиля

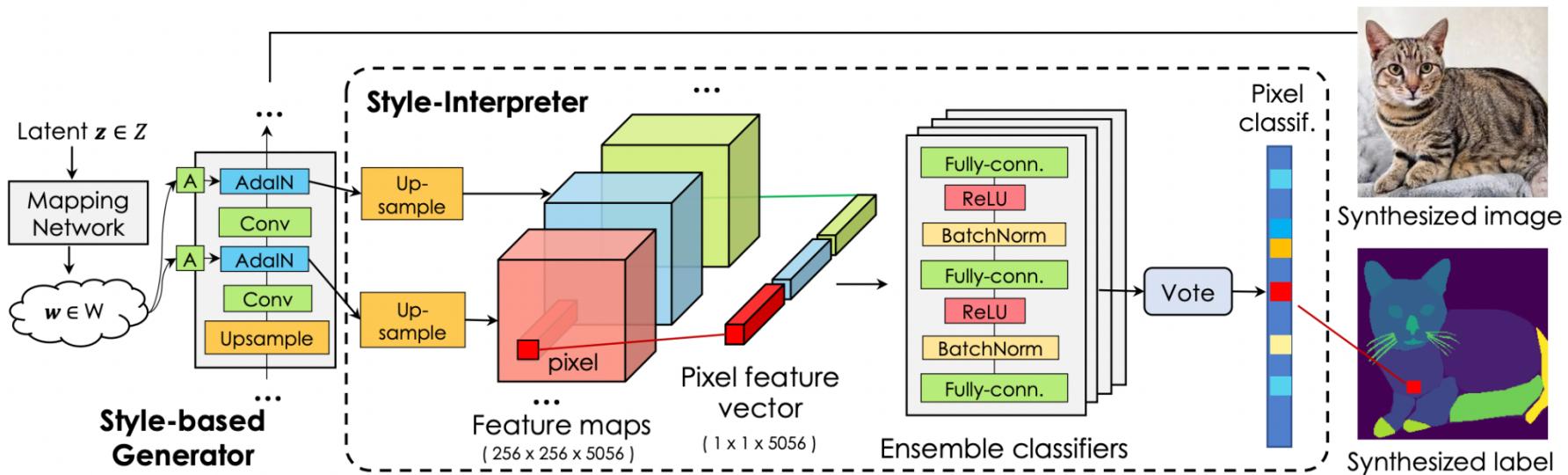
Стиль задается парой векторов  $y_s$  и  $y_b$ . Их размерность совпадает с размерностью  $x$ . Применение стиля делается с помощью AdaIN

$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i},$$

# Процесс обучения

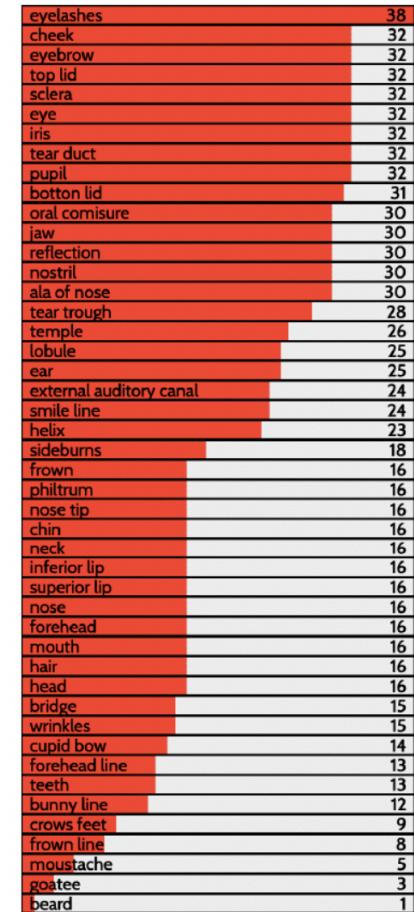
Учим как обычный прогрессивно растущий GAN. То есть сначала учимся генерировать изображение 4x4, потом 8x8, и так далее, добавляя всё больше слоёв.

# Архитектура интерпретатора

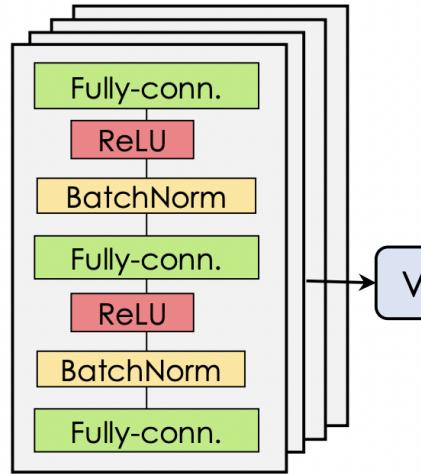


# Оптимизация обучения

Мы не можем за раз обучить модель на всех признаках для всех пикселей, потому что размерность слишком большая ( $256 \times 256 \times 5056$  или больше), поэтому выбираем случайные пиксели. При этом следим, чтобы каждый из размеченных классов обязательно был представлен



# Классификаторы



Ensemble classifiers

Попиксельные классификаторы это просто полносвязные сетки, которые используют одни и те же веса для каждого пикселя

Используем ансамбль классификаторов, чтобы сгладить шум

Исходя из задачи выбираем лосс и способ работы с ансамблем

## Использование для генерации датасета

Генерируем картинку с помощью генератора, генерируем разметку с помощью интерпретатора

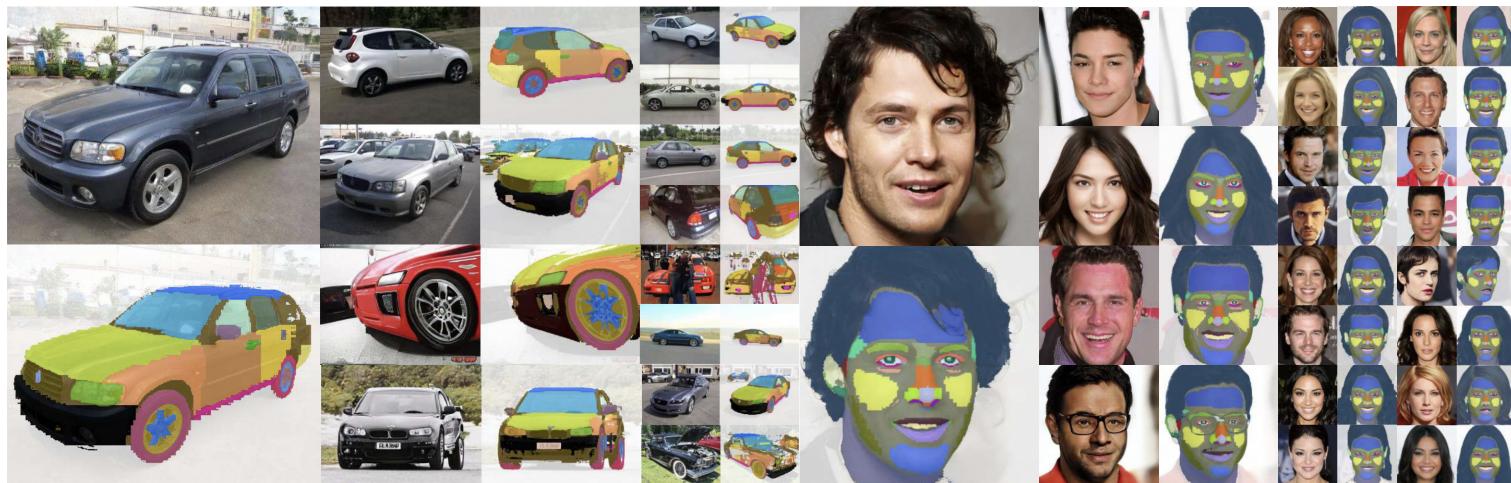
Иногда StyleGAN может генерировать артефакты. Дискриминатор не всегда их замечает. Вместо него пытаемся найти шум с помощью ансамбля в интерпретаторе. Смотрим, насколько разные классы дают отдельные классификаторы

Оставляем в выборке только те изображения, на которых метрика не слишком плохая, в статье это лучшие 90%

# Преимущества подхода

Можно сгенерировать очень большой датасет

Для обучения интерпретатора нужно разметить очень небольшое число синтетических картинок, в статье это десятки изображений



# Недостатки подхода

На сгенерированных изображениях может быть меньше деталей

Разметка может получиться хуже



Рецензент

# Сильные и слабые стороны

- Актуальность
- Новизна
- Плохое описание некоторых алгоритмов
- Применимость

Исследователь

# Общая информация о статье

- Статья 2021 года
- Представлена на CVPR 2021 в качестве Oral paper

# Авторы

Всего 8 авторов, расскажу про первого в списке.

- У автора 3 статей, все про CV.
- Самая похожая на статью из доклада вышла за год неё.
- В ней автор с помощью GAN синтезировал изображения трёхмерных объектов с разных ракурсов

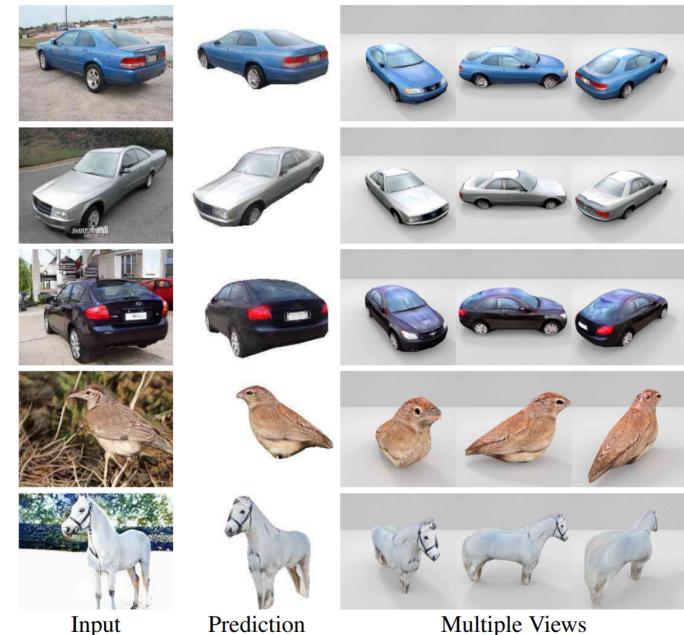


Yuxuan Zhang

[Princeton University](#)

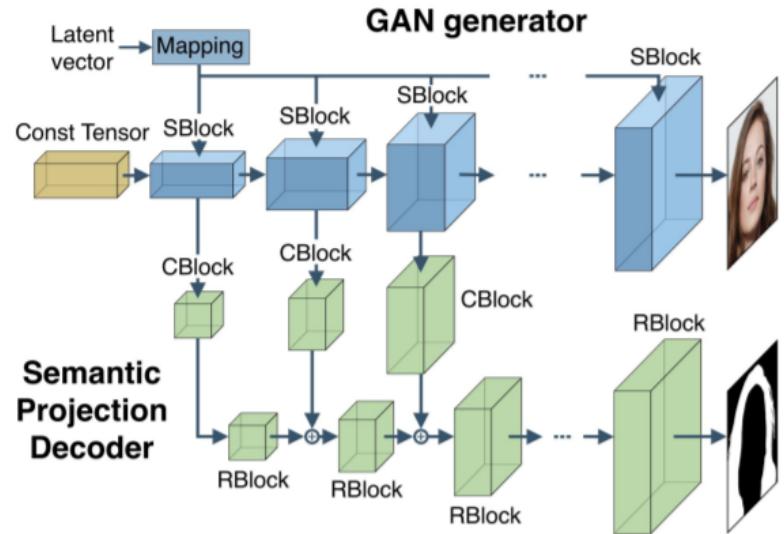
Verified email at princeton.edu

Machine learning Computer Vision



# Конкуренты

- На момент публикации уже была очень похожая работа, авторы указали её в related work.
- Вместо ансамбля простых классификаторов тут используют свёртки.
- Авторы утверждают, что их подход лучше (доказательств не нашёл)

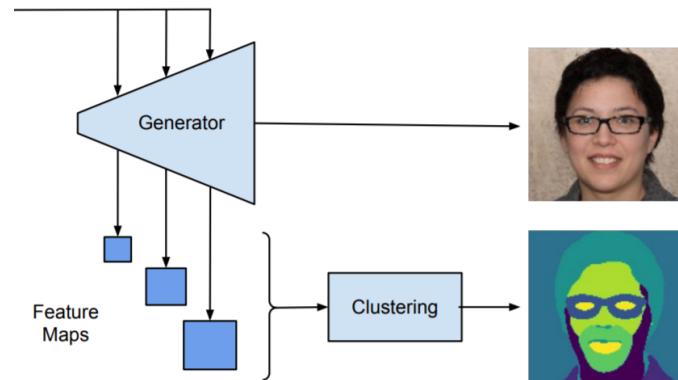


<https://arxiv.org/pdf/2006.10451.pdf>

# Ссылки на работу (целых 6 штучек)

- В основном статью вставляют в related work, когда синтезируют что-то с помощью GAN
- Есть статья, в которой авторы с помощью GAN делают сегментацию, но unsupervised. Они заявляют, что их подход превосходит результаты статьи (нет)

<https://arxiv.org/pdf/2107.12518.pdf>



Хакер

# Как обучить DatasetGAN?

- пример с дефолтной задачей
- пример с изменением классов в дефолтной задаче
- эту модель так просто не обучить
- V100 в яндекс облаке с 64гб оперативной памяти учится ~6 часов
- для стайлгана этого мало

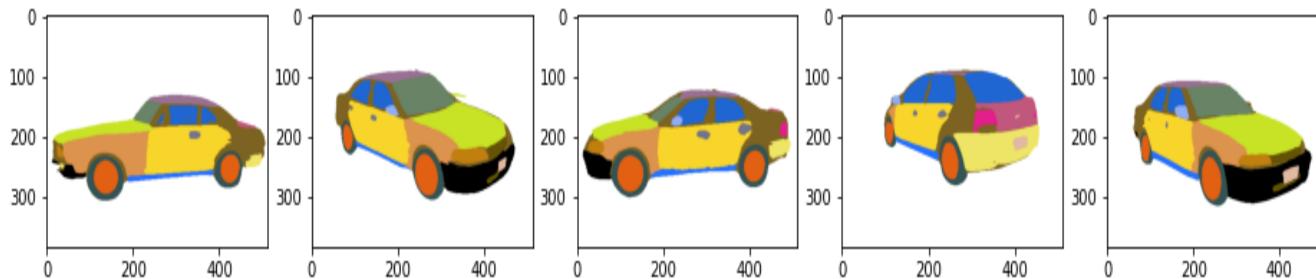
# Как обучить DatasetGAN?

- datasetGAN/experiments - конфиги экспериментов
- [https://github.com/nv-tlabs/datasetGAN\\_release/blob/d9564d4d2f338eaad78132192b865b6cc1e26cac/datasetGAN/train\\_interpreter.py#L130](https://github.com/nv-tlabs/datasetGAN_release/blob/d9564d4d2f338eaad78132192b865b6cc1e26cac/datasetGAN/train_interpreter.py#L130) - выставить параметры нужного StyleGAN
- utils/data\_util.py - добавить свои классы и палитру

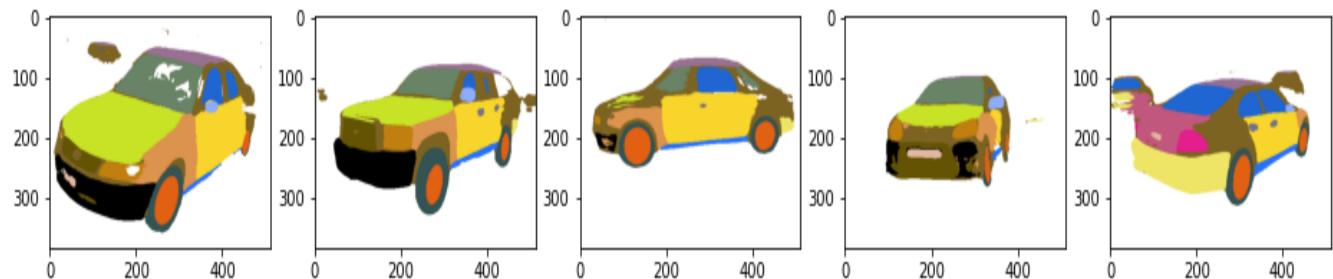
```
{  
    "exp_dir": "model_dir/car_29",  
    "batch_size": 32,  
    "category": "car_29",  
    "debug": true,  
    "dim": [384, 512, 5056],  
    "deeplab_res": 512,  
    "number_class": 29,  
    "testing_data_number_class": 20,  
    "max_training": 16,  
    "stylegan_ver": "1",  
    "annotation_data_from_w": false,  
    "annotation_mask_path": "./synthetic_train_data/images_masks_29",  
    "testing_path": "./synthetic_train_data/images_masks_29",  
    "average_latent": "./dataset_release/training_latent/car_20/avg_latent_stylegan1.npy",  
    "annotation_image_latent_path": "./dataset_release/training_latent/car_20/latent_stylegan1.npy",  
    "stylegan_checkpoint": "./dataset_release/stylegan_pretrain/karras2019stylegan-cars-512x384.for_g_all.pt",  
    "model_num": 10,  
    "upsample_mode": "bilinear"  
}
```

# Эксперимент с дефолтной задачей

Качественные маски:



Грязные маски:

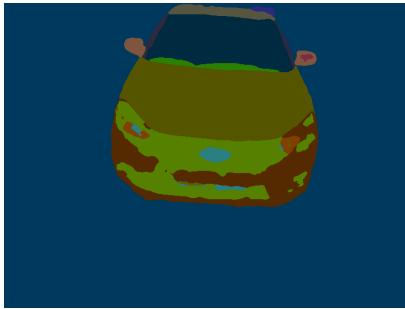


# Эксперимент с дефолтной задачей

- добавили синтетические примеры в выборку (1000 реальных, 1000 синтетических)
- качество на нашей выборке немного упало, однако модель узнала больше о внешнем мире

# Эксперимент с дефолтной задачей

Без синтетики



Оригинал



С добавлением синтетики



# Эксперимент с другим набором классов

- разметили примеры авторов своими классами (29 вместо 20, добавили мелкие детали)
- мелкие детали не сгенерировались адекватно
- маски стали более грязными
- модель все еще сохраняла способность угадывать класс региона при очень близком ракурсе

# Эксперимент с другим набором классов



# Выводы

- маски часто выходят некачественными, надо фильтровать
- выдает хороший результат, если классы сбалансированы и присутствуют на всех примерах
- синтетическая выборка полезна для модели, потому что учит ее воспринимать случаи, которые она раньше не видела
- более тонкие настройки - внутри StyleGAN