

SinGAN: Learning a Generative Model from a Single Natural Image

Валитов Эльдар, БПМИ-172

Суть

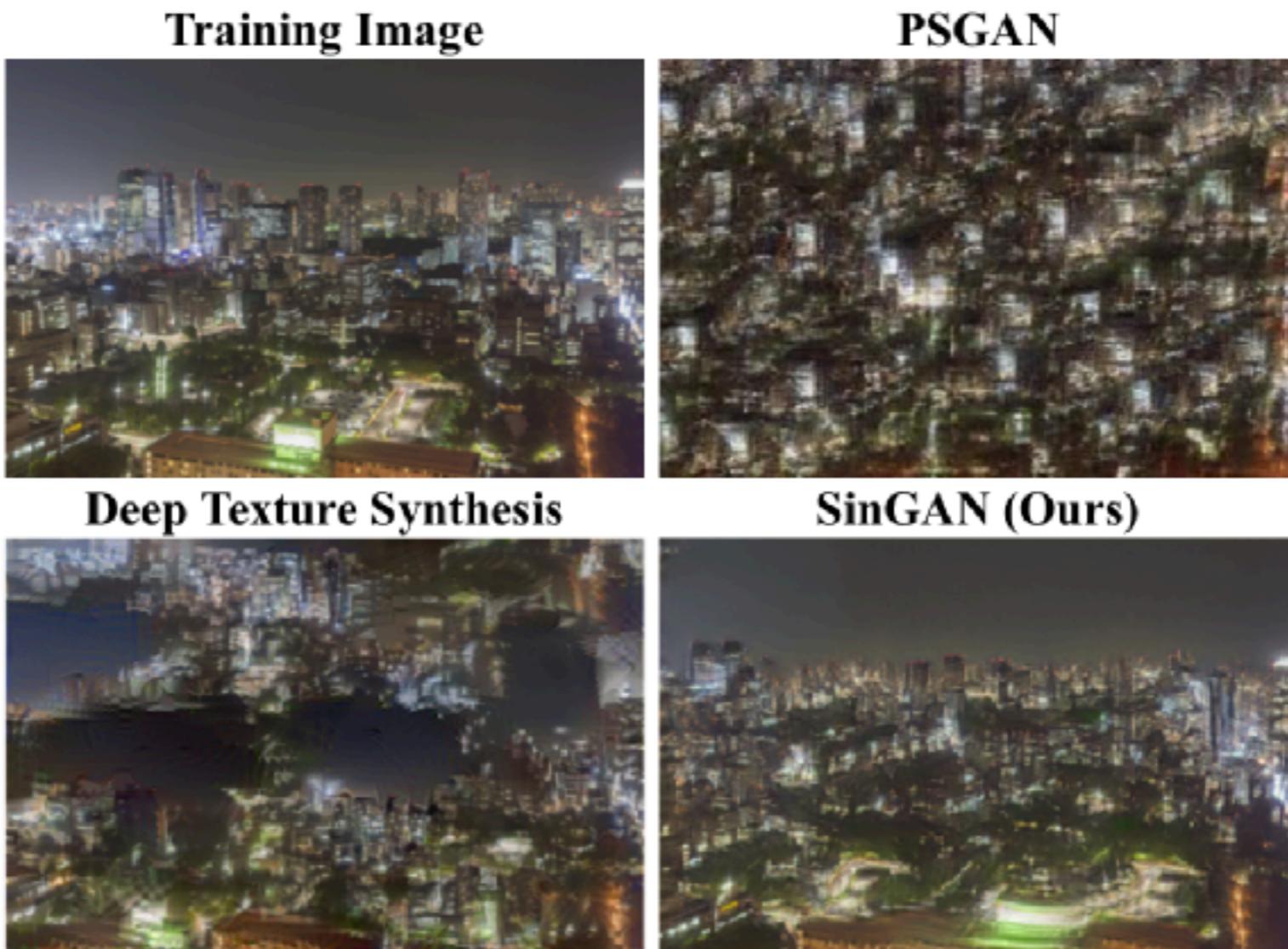
Single training image



Мотивация к созданию. Что было до?

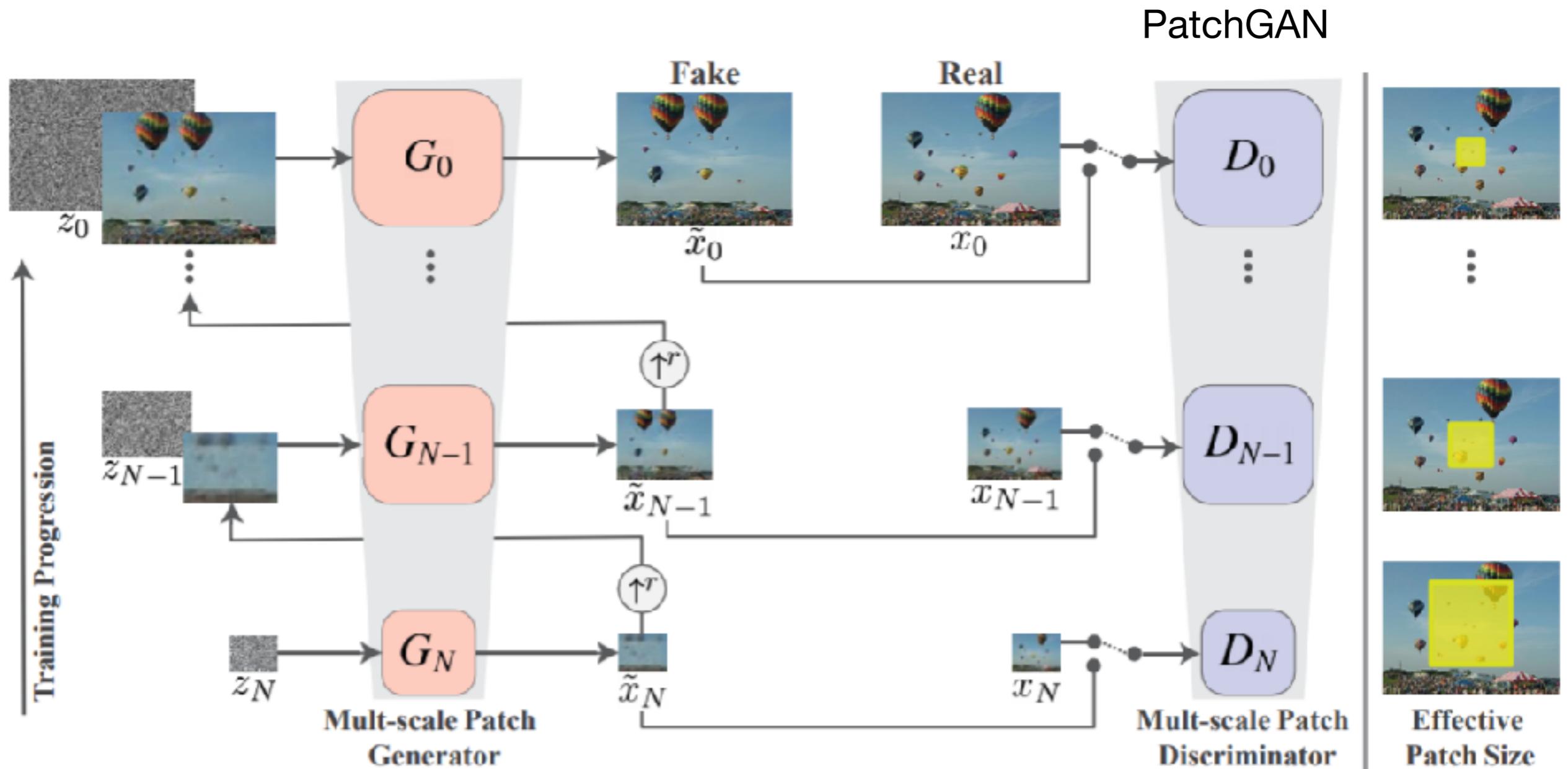
- Попытки переобучить модель на одном изображении, чтобы решить конкретную задачу (как пример - Deep Prior)
- InGAN - схожая идея как у SinGAN, однако сеть является условной, в то время как в SinGAN условия не требуются
- Были GAN с multi-scale архитектурой, но не подстроенной под обучение на одном изображении

Мотивация к созданию



SinGAN vs Single Image Texture Generation

Структура



x_n - downsampled изображение для определенного шага

Генерации

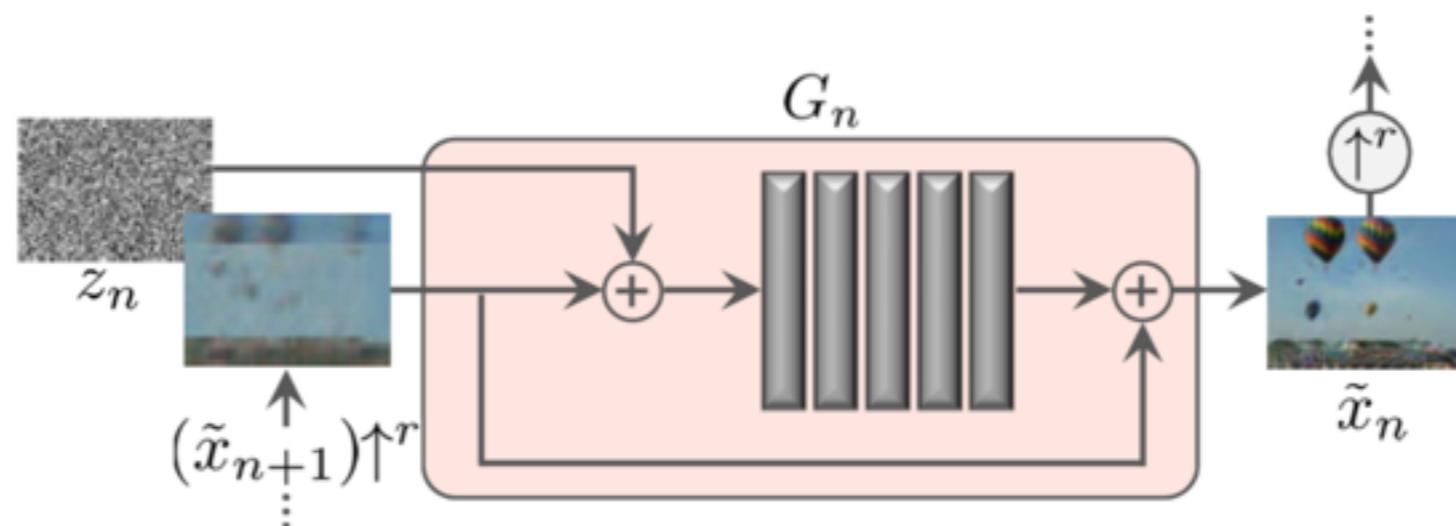
$\{G_0, \dots, G_N\}$ - пирамида генераторов

$\{x_0, \dots, x_N\}$ - изображения

$\tilde{x}_N = G_N(z_N)$ - генерация на 1ом шаге

$\tilde{x}_n = G_n(z_n, (\tilde{x}_{n+1})^{\uparrow r}), \quad n < N.$ - генерация на каждом последующем

$\tilde{x}_n = (\tilde{x}_{n+1})^{\uparrow r} + \psi_n(z_n + (\tilde{x}_{n+1})^{\uparrow r})$ - раскрыли генератор, фи - конволовионная сеть



ФУНКЦИЯ ПОТЕРЬ

$$\min_{G_n} \max_{D_n} \mathcal{L}_{\text{adv}}(G_n, D_n) + \alpha \mathcal{L}_{\text{rec}}(G_n)$$

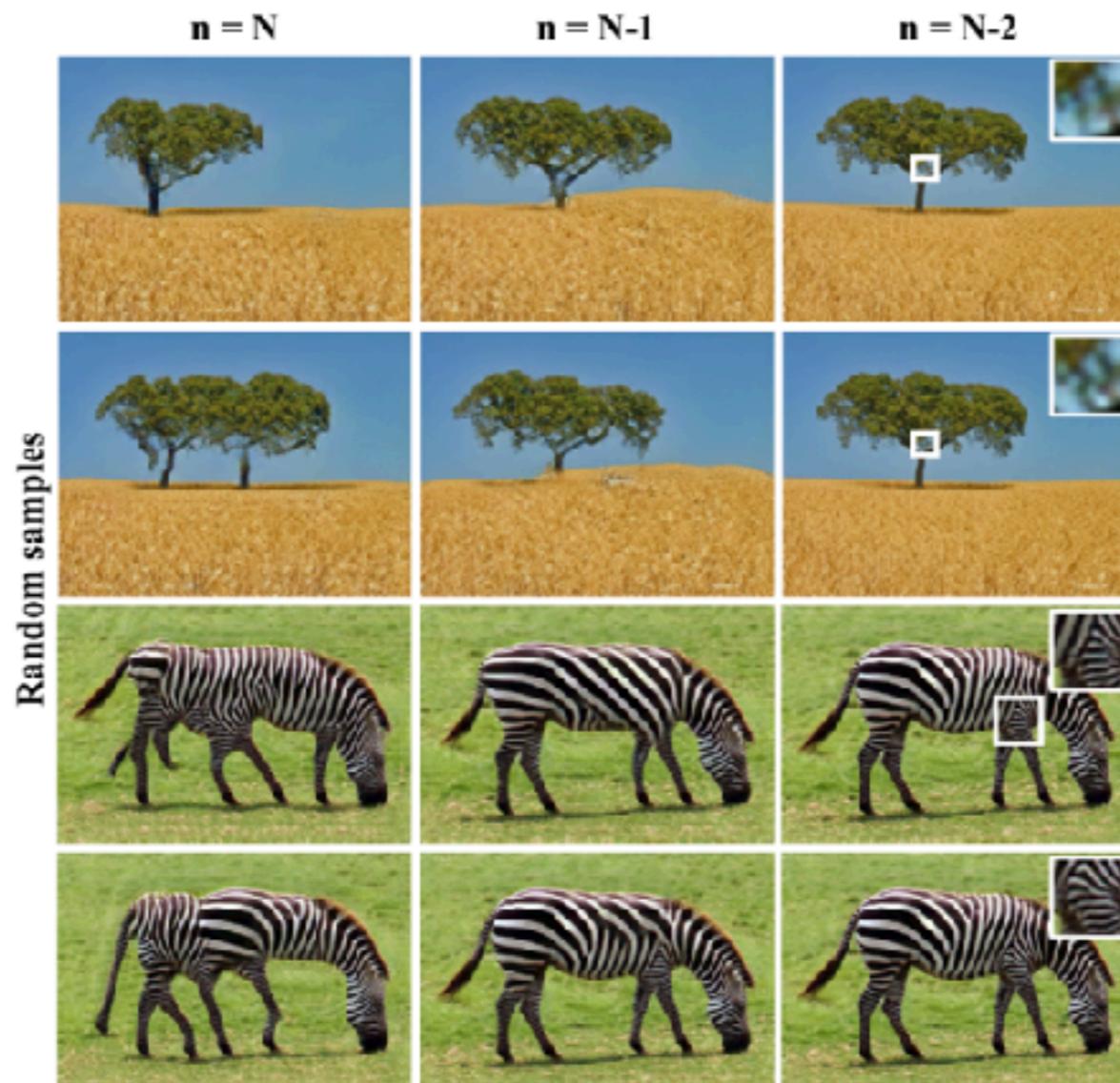
Adversarial loss (L_adv) - WGAN GP loss

$$\mathcal{L}_{\text{rec}} = \|G_n(0, (\tilde{x}_{n+1}^{\text{rec}})^{\uparrow r}) - x_n\|^2, \quad \text{for } n < N$$

$$\mathcal{L}_{\text{rec}} = \|G_N(z^*) - x_N\|^2 \quad \text{for } n = N$$

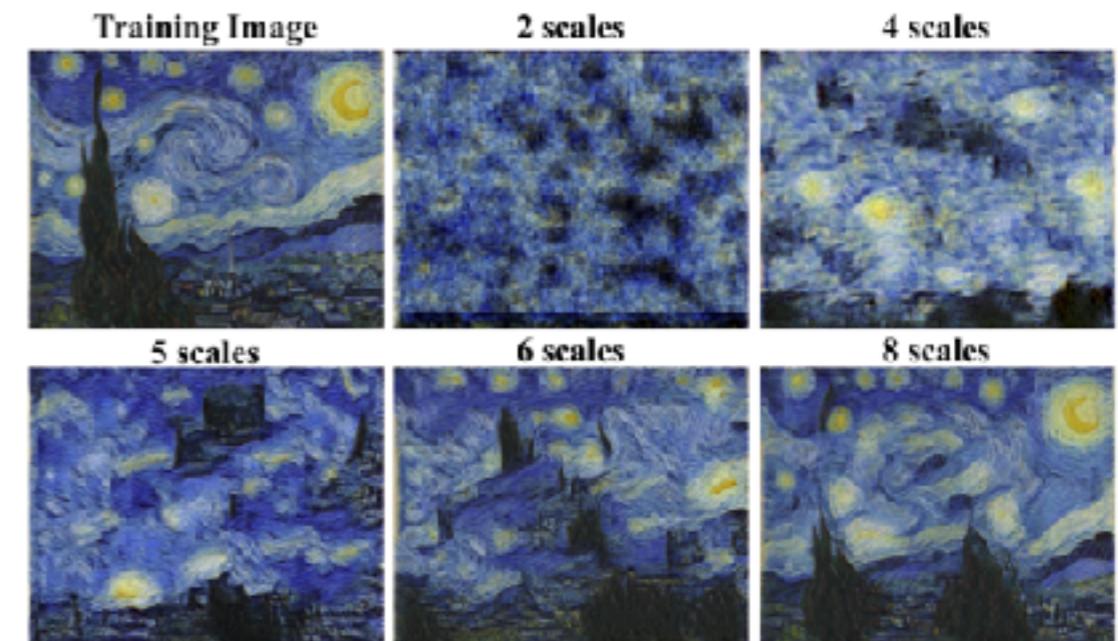
Важный момент

- Поскольку SinGAN дает возможность подавать на вход изображение на произвольном масштабе, можно таким образом влиять на итог генерации



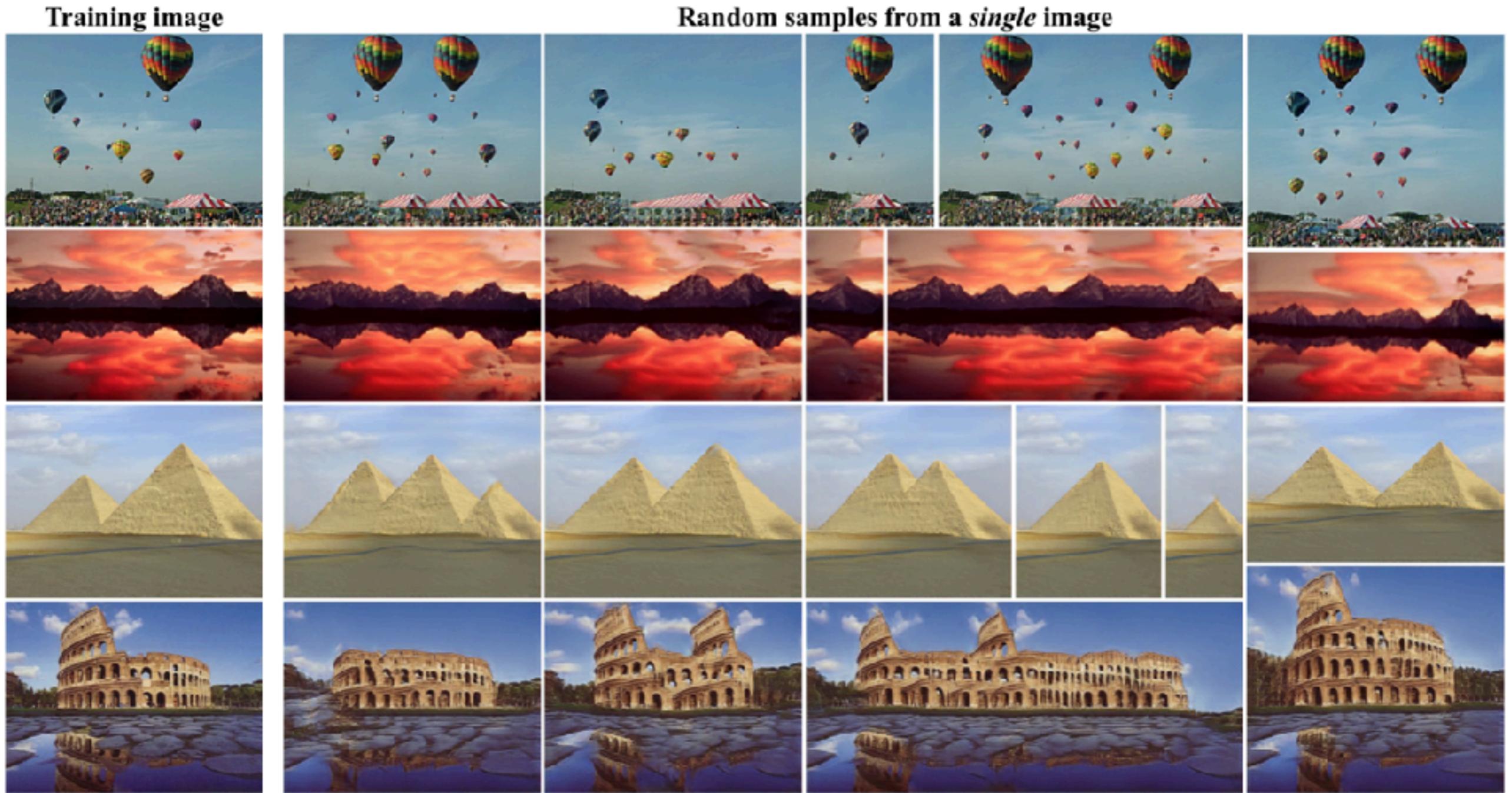
Важный момент

- При маленьком количестве шагов модель уловит только основную структуру изображения, при большом - модель учится понимать структуры большего размера и их взаимодействие
- По проведенным тестам, люди не всегда справляются с определением сгенерированного SinGAN изображения, при этом еще раз можно заметить, что шаг на котором мы подаем сети изображение сильно влияет на результат



1st Scale	Diversity	Survey	Confusion
N	0.5	paired	$21.45\% \pm 1.5\%$
		unpaired	$42.9\% \pm 0.9\%$
$N - 1$	0.35	paired	$30.45\% \pm 1.5\%$
		unpaired	$47.04\% \pm 0.8\%$

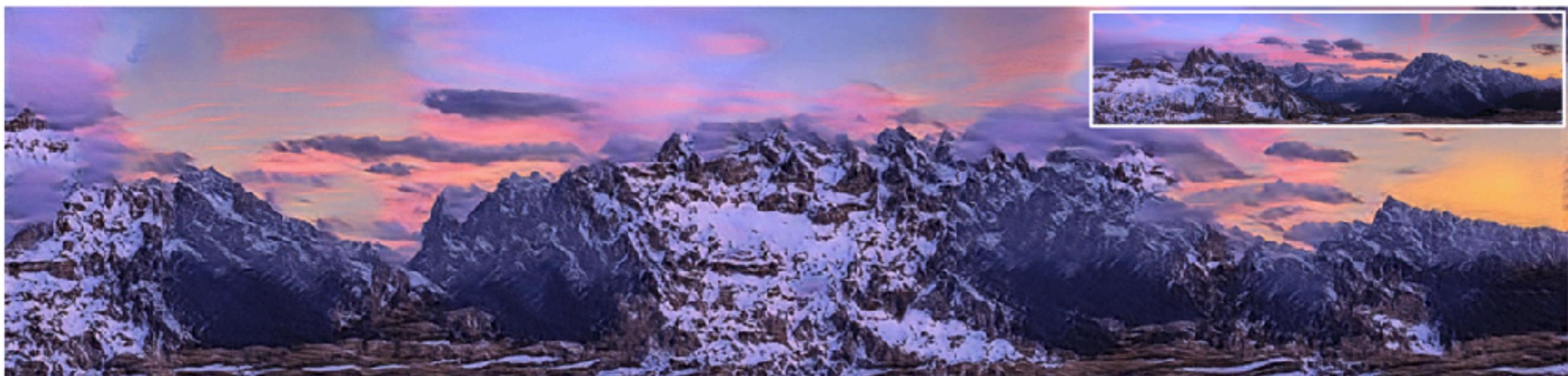
Примеры работы



Примеры работы

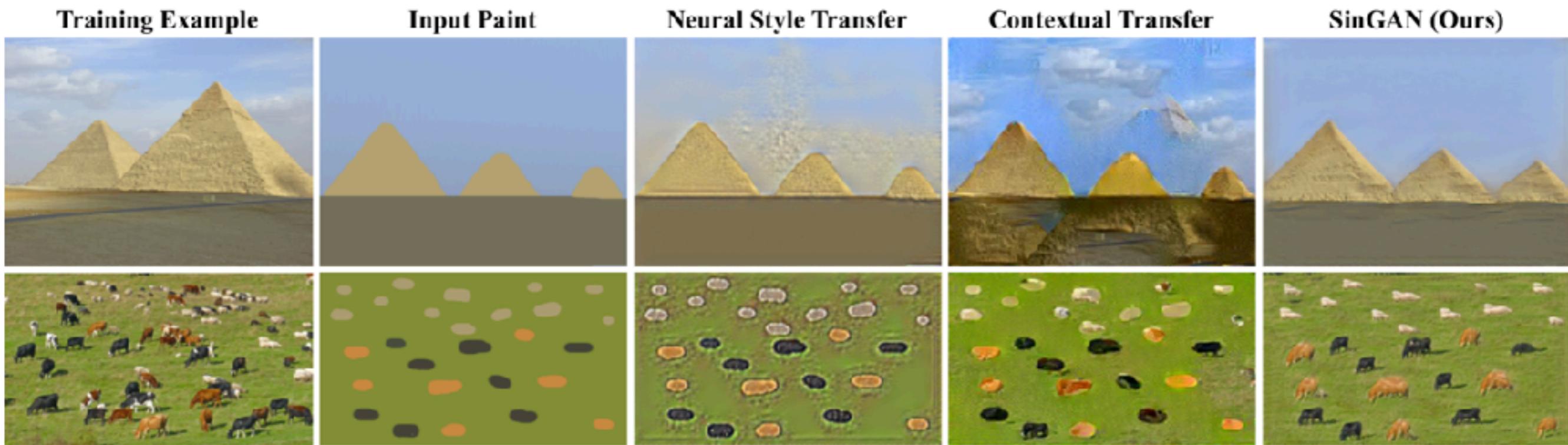
Super-Resolution - увеличение изображения размера a в r раз. Для этого берем $\alpha = 100$ в части функции потерь с reconstruction loss и изменение масштаба равное $r = \sqrt[k]{s}$.

Далее мы увеличиваем размер в r раз и подаем изображение в G_0 . Так повторяем k раз

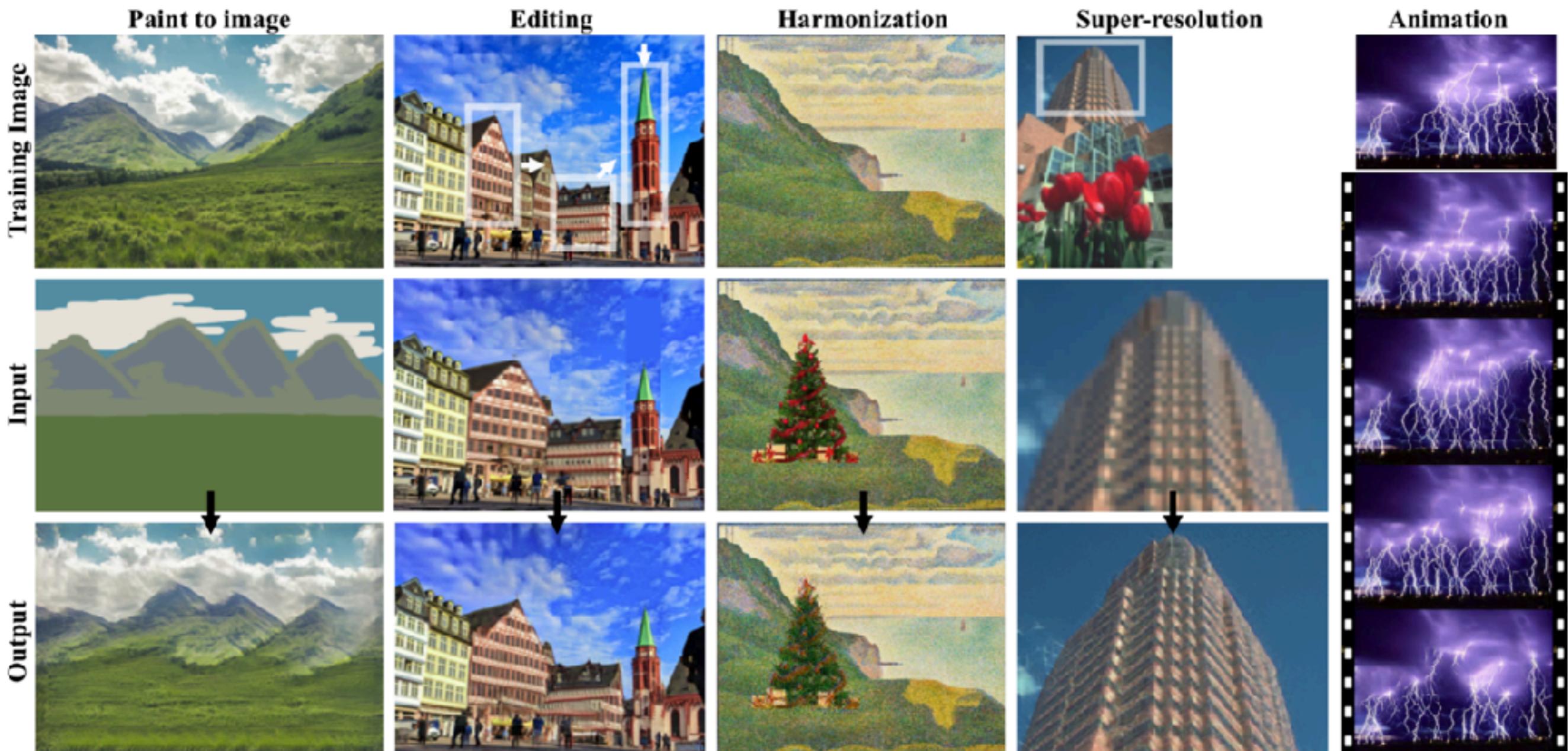


Примеры работы

Paint to image - производим downsampling (понижение качества) изображения и подаем на вход в один из нижних (например N-1, N-2) слоев. Таким образом, сгенерированное изображение будет иметь схожую структуру, сохранив реалистичную структуры и частотные характеристики



Примеры работы



Заключение

- SinGAN - мощный инструмент в области CV с оригинальными решениями в структуре (до этого не очень популярно было применять остаточное обучение), и хорошими результатами.

А что дальше?

- Одним из очевидных недостатков SinGAN является то, что все сгенерированные изображения будут принадлежать плюс-минус одному классу, дальнейшие исследования скорее всего будут направлены на создание GAN, которая будет обучаться на одном изображении и генерировать выходы разных классов

Источники

- <https://arxiv.org/pdf/1905.01164.pdf> - статья
- <https://github.com/tamarott/SinGAN> - репозиторий с кодом
- <https://www.youtube.com/watch?v=j9id4UpN9BI> - доклад по SinGAN от его авторов, проясняющий всю суть