

# StyleGAN2 by NVIDIA

Генеративно-состязательные сети (Generative Adversarial Networks — **GAN**) — это принцип обучения генеративных моделей с помощью двух нейронных сетей. А именно, генератор, который отображает случайный вектор в изображение, и дискриминатор, который пытается отличить сгенерированное изображение от настоящего. После обучения обычно используется только генератор для генерации синтетических изображений. До выхода этой статьи, state-of-the-art архитектурой в безусловном (unconditional) генеративном моделировании изображений в высоком разрешении (1024x1024) на основе данных была модель StyleGAN от NVIDIA. Однако, многие исследователи заметили каплевидные артефакты в сгенерированных изображениях. А также артефакты, связанные с прогрессивным ростом (progressive growing), который оказался очень успешным в стабилизации GAN высокого разрешения.

## Замена Adaptive Instance Normalization

Исследователи установили, что проблема связана с операцией **AdaIN**, которая нормализует среднее значение и дисперсию каждой feature map по отдельности, тем самым потенциально уничтожая любую информацию, найденную в величинах признаков относительно друг друга. Также их гипотеза подтверждается тем, что, когда этап нормализации удаляется из генератора, каплевидные артефакты полностью исчезают.

Поэтому было предложено заменить **AdaIN** на **modulated convolutions**. А именно, устроены они так:

$$w'_{ijk} = s_i \cdot w_{ijk}$$

Где  $w$  и  $w'$  исходные и модулированные веса соответственно. А  $s_i$  — это масштаб, соответствующий  $i$ -й входной карте признаков, а  $j$  и  $k$  перечисляют выходные карты признаков и пространственный след свертки соответственно.

Далее, для того, чтобы вернуть выходные значения до стандартного отклонения, применяется стандартная нормализация, а именно:

$$w''_{ijk} = w'_{ijk} / \sqrt{\sum_{i,k} w'^2_{ijk} + \epsilon}$$

Где  $\epsilon$  — маленькая константа, для избежания числовых проблем.

### Path length regularization

Также исследователи заметили корреляцию между воспринимаемым качеством изображения и perceptual path length, метрикой, которая была первоначально введена для количественной оценки гладкости отображения скрытого пространства на выходное изображение путем измерения средних расстояний LPIPS между сгенерированными изображениями при небольших пертурбациях в латентном пространстве. Опять же, меньший PPL (более гладкое отображение генератора), по-видимому, коррелирует с более высоким общим качеством изображения, в то время как другие показатели не видят изменений.

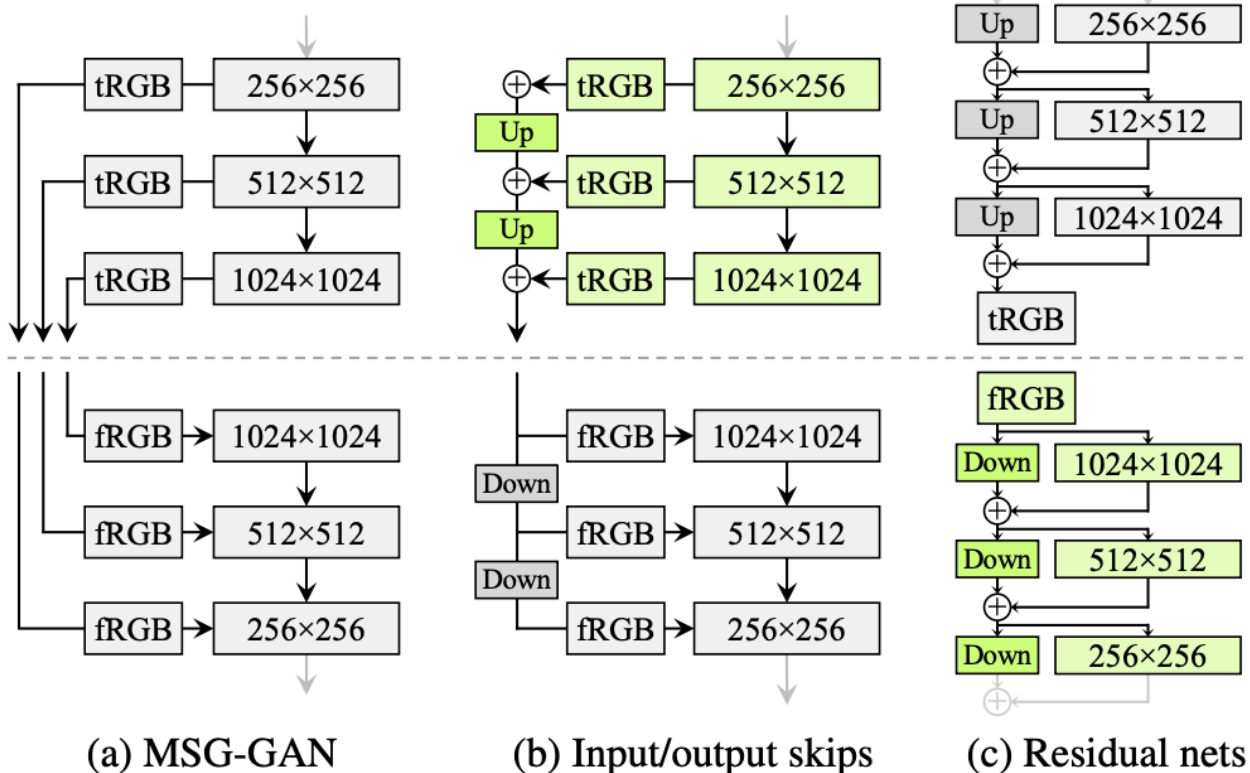
Поэтому авторами статьи был предложен новый регуляризатор (Path length regularization):

$$\mathbb{E}_{\mathbf{w}, \mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})} \left( \left\| \mathbf{J}_{\mathbf{w}}^T \mathbf{y} \right\|_2 - a \right)^2$$

Где  $\mathbf{J}_{\mathbf{w}}$  — матрица Якоби генератора,  $\mathbf{y}$  — случайные изображения с нормально распределенными интенсивностями пикселей, а  $\mathbf{w} \sim f(\mathbf{z})$  ( $f$  — это mapping network), где  $\mathbf{z}$  — нормально распределены.

### Progressive growing revisited

Также авторы статьи пересмотрели архитектуру генератора и дискриминатора. Были рассмотрены следующие варианты (рисунок):



Исходя из полученных метрик (FID и PPL) на двух датасетах (FFHQ и LSUN Car) было замечено следующее: skip connections в генераторе значительно улучшает PPL во всех конфигурациях, а residual дискриминатор явно повышает FID.

## Вывод

Авторы выявили и исправили несколько проблем с качеством изображения в StyleGAN, что привело к дальнейшему повышению качества и значительному усовершенствованию некоторых наборов данных. Производительность тренировок также улучшилась. Это удивительно, но авторы даже написали свои кастомные CUDA-кernels, которые ускоряют тренировку на 30%.