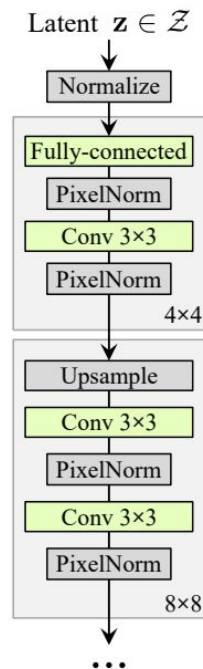


StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets

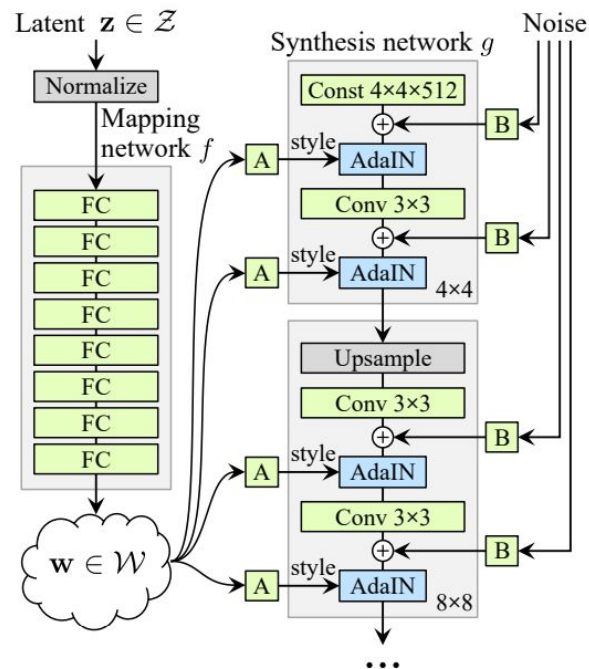
Докладчик: Артем Макоян, БПМИ192

StyleGAN

- Вход z сопоставляем промежуточному коду латентного пространства W и получаем w
- К w применяем обученное аффинное преобразование
- Оно контролирует блок AdaIN на каждом слое
- Перед каждой свертки подмешиваем шум
- Шум перед этим масштабируем поканально



(a) Traditional

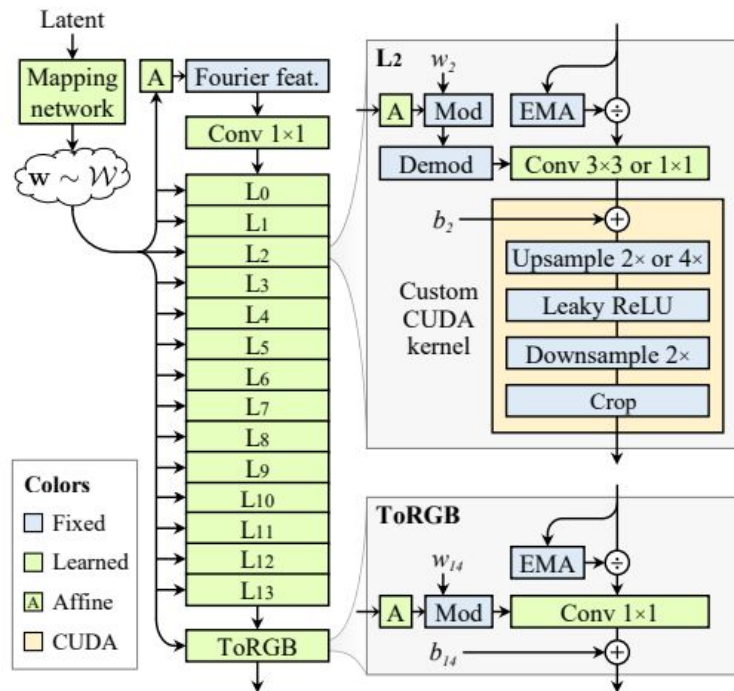


(b) Style-based generator

StyleGAN3

Хотим сделать StyleGAN эквивариантный сдвигам и поворотам (или как можно ближе к эквивариантному)

- Заменяем константный вход на фичи Фурье
- Убираем попиксельный шум
- Уменьшаем глубину и убираем skip connections у выхода
- Заменяем upsample на Kaiser window
- Перед нелинейностью ставим m upsample-ов, а после m downsample-ов (анти алиасинг)
- и много других трюков...



Проблема StyleGAN-ов

- StyleGAN сложно учить
- StyleGAN хорош на маленьких доменах, но когда картинки очень разнообразны и их много - он работает плохо

ProjectedGAN

$$\min_G \max_D \left(\mathbb{E}_{\mathbf{x}}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}}[\log(1 - D(G(\mathbf{z}))) \right]$$

$$\min_G \max_{\{D_l\}} \sum_{l \in \mathcal{L}} \left(\mathbb{E}_{\mathbf{x}}[\log D_l(P_l(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}}[\log(1 - D_l(P_l(G(\mathbf{z})))) \right]$$

Теперь у нас много независимых дискриминаторов и на вход им подаются не картинки, а их проекции (проекции необучаемые). Проекции можем получать из предобученных моделей, случайные проекции и т.д.

Повышает скорость сходимости и качество, делает обучение более устойчивым.

StyleGAN-XL

Если кратко - это StyleGAN3, который обучили следуя идеям ProjectedGAN.

- Регуляризации вводились для того, чтобы распределение картинок было больше унимодальным, но теперь это только вредит.
- Но регуляризация путей позволяет делать высококачественную инверсию
- Решение: начинаем использовать регуляризацию путей только после того, как модель много тренировалась (200к картинок)
- Для дискриминатора используем спектральную нормализацию без штрафа за градиенты
- Первые шаги обучения блёрим первые 200к картинок. Это позволяет сосредоточиться дискриминатору на глобальных фичах

StyleGAN-XL

- “Скрытая размерность” датасета картинок довольно маленькая и сильно меньше классического StyleGAN (512).
- Из-за такого “гэпа” генератору сложнее учиться особенно на начальных шагах
- Поэтому меняем размерность z с 512 на 64

StyleGAN-XL. Условные GAN-ы.

- Обуславливание GANов информацией о GANах контролируют классы + улучшают общее качество генерации
- Каждый класс представляется эмбедингом и конкатенируется с z
- Оказалось, что в случае обучения в стиле ProjectedGAN распределение классов вырождается (recall очень низкий)
- Вместо простого обучения эмбедингов с нуля возьмем предобученные эмбединги полученные усреднением фичей Efficientnet-lite0, к которым применили линейную проекцию (нужна размерность Z)

StyleGAN-XL. Прогрессивный рост генератора.

- Хочется не сразу обучаться на высоком разрешении, а двигаться от низкого разрешения к высокому
- Сначала берем обычный StyleGAN и обучаем его на низком разрешении, при шаге повышения размерности - удаляем последние 2 слоя у генератора и добавляем 7 новых. Тем самым мы постепенно увеличиваем мощность модели
- При такой тактике возникает проблема анти-алиасинга
- StyleGAN3 отлично борется с антиалиасингом, надо просто правильно использовать его слои в прогрессивном росте генератора

StyleGAN-XL. Использование нескольких предобученных сетей

- Какие предобученные сети стоит использовать для проецирования картинок в ProjectedGAN?
- Исследования показывают, что особой разницы нету, но что насчет комбинирования?
- Наилучший результат был получен смешиванием EfficientNet и ViT, причем довольно сильное улучшение

StyleGAN-XL. Влияние классификатора на ГАНЫ

- Изначально использовали для диффузионок - на шаге t добавляли градиент предобученного классификатора
- Пропускаем сгенерированную картинку через классификатор и берем CE loss
- Таким образом лучше генерируем картинки относительно классов и повышаем разнообразие сгенерированных датасетов



Fig. 3. **Samples at Different Resolutions Using the Same w.** The samples are generated by the models obtained during progressive growing. We upsample all images to 1024^2 using nearest-neighbor interpolation for visualization purposes. Zooming in is recommended.