

# StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets (Axel Sauer, Katja Schwarz, Andreas Geiger)

Автор обзора-рецензии: Николаев Максим, 192

*1. Опишите суть работы в паре предложений, выделите ее основной вклад.*

Архитектура StyleGAN имеет множество хороших свойств, но плохо работает на больших и разнообразных наборах данных, поэтому авторы в своей работе постарались переработать принцип обучения StyleGAN 3 и перенесли на него множество парадигм из других работ, чтобы научить его хорошо генерировать любые домены, не потеряв свойства архитектуры.

*2. Когда написана работа? Опубликована ли она на какой-то конференции? Кто ее авторы, есть ли у них другие схожие работы? Подумайте как авторы пришли к идее статьи?*

Первая версия статьи вышла 1 февраля 2022 года, а последняя версия 5 мая 2022 года. Статья попала на конференцию SIGGRAPH 22. Авторы статьи: Axel Sauer, Katja Schwarz – оба PhD студенты University of Tübingen, а также их преподаватель Andreas Geiger профессор компьютерных наук.

Axel Sauer и Andreas Geiger за 3 месяца до StyleGAN-XL, 1 ноября 2021 года, опубликовали статью Projected GANs Converge Faster, в которой протестировали множество парадигм для GANs и которые после попадут в StyleGAN-XL. Их основное достижение – стабилизация и ускорение обучения, статья попала на NeurIPS 2021. Andreas Geiger в частности имеет множество статей по 3D computer vision, беспилотникам и GANs.

Основной идеей для статьи послужил Projected GANs, после которого они решили попробовать перенести множество техник на архитектуру StyleGAN.

*3. Какие из статей в списке ссылок оказали наибольшее влияние на данную работу? Можно ли выделить какие-то 1-3 статьи, которые можно назвать базовыми для этой работы?*

Основные 2 работы: Alias-Free Generative Adversarial Networks (StyleGAN3) – их базовая модель, которую они последовательно улучшали и Projected GANs Converge Faster – откуда они брали основные парадигмы.

StyleGAN3 – Модель из семейства StyleGAN, в которой значительно переработали и немного урезали архитектуру StyleGAN2, чтобы исправить проблемы при генерации видео в их латентном пространстве, в частности исправлены проблемы с привязкой волос к координатам.

Projected GANs Converge Faster – В этой статье предлагается использовать несколько дискриминаторов, в которые картинки попадают после применения

некоторой проекции, таким образом получается ансамбль из разных дискриминаторов, также авторы в этой работе пересмотрели принцип применения регуляризации и показали, что для таких наборов данных как ImageNet латентное пространство сильно меньше, чем латентное пространство в популярных GAN архитектур и после его уменьшения показали стабилизацию обучения.

***4. Кто цитирует данную статью? Есть ли у этой работы прямые продолжения, которые стоит прочесть тем, кто заинтересовался этой работой?***

Прямого продолжения у работы пока нет, цитат у данной работы на момент написания рецензии (17.11.2022) – 46, в основном цитирование статьи идёт в работах связанном с редактированием изображений в латентном пространстве StyleGAN, а также наткнулся на интересную статью “The Role of ImageNet Classes in Fréchet Inception Distance” от авторов из Nvidia, в частности от самого автора StyleGAN, в которой они подробно исследуют FID метрику и в частности показывают, что при лучшем FID у модели Projected GANs по сравнению со StyleGAN 2, последний генерирует более качественные картинки для человеческого восприятия без артефактов, которые мы также наблюдали у StyleGAN-XL в их статье.

***5. Есть ли у работы прямые конкуренты? Опишите как соотносится данная работа с этими конкурентами.***

Основной конкурент среди GANs был BigGAN, считавшийся SOTA архитектурой для генерации больших и разнообразных наборов данных по типу ImageNet и которого до этого StyleGAN не мог обогнать по качеству на них. Также среди конкурентов есть CDM и ADM – диффузионные модели, но на всех основных метриках FID и IS – StyleGAN-XL оказался лучше, но всё ещё проигрывает им по precision и recall при conditional генерации. В целом для генерации картинок большого разрешения 512+ у StyleGAN-XL нет конкурентов, а для картинок маленького разрешения появилась диффузионная модель EDM, которая обогнала StyleGAN-XL на CIFAR-10 и ImageNet, а также StyleNAT – GAN на основе трансформера, который смог обогнать StyleGAN-XL при генерации на FFHQ 256.

***6. Опишите сильные, на ваш взгляд, стороны работы. Стоит обратить внимание на корректность утверждений в работе, значимость и новизну вклада, актуальность для исследовательского сообщества, понятность текста и воспроизводимость результатов.***

Сейчас StyleGAN-XL SOTA для генерации высокоразмерных наборов данных и авторы внесли большой вклад в развитие GANs утвердив и доказав работоспособность множества парадигм. Оцениваю актуальность и значимость для исследовательского сообщества как высокую. Текст работы был понятен, а настройка репозитория для генерации изображений проходит легко.

Также из плюсов работы хочется отметить скорость инференса, по сравнению с диффузионками, а также распутанное латентное пространство, которое можно использовать для редактирования изображений.

***7. Опишите слабые, на ваш взгляд, стороны работы, обращая внимание на те же моменты, что и в предыдущем пункте.***

Основные минусы: модель слишком большая по сравнению с остальными GANs, также можешь всё ещё не очень стабильна и из-за этого её приходится обучать прогрессивно, по качеству проигрывает диффузионкам на маленьких разрешениях, всё ещё проигрывает другим GANs и диффузионкам в точности и разнообразии генерации на больших наборах данных.

***8. Предложите как можно было бы улучшить статью: какие дополнительные утверждения/эксперименты стоило бы рассмотреть, какие вопросы остались не закрытыми для вас после прочтения статьи, обсуждение связи с какими работами дополнило бы работу.***

Лично у меня возникают вопросы к настоящему качеству генерации модели, потому что при тестировании и даже на картинках в самой статье часто встречается множество артефактов по сравнению с другими работами, поэтому в качестве дальнейшей работы я бы предложил переизмерить качество генерации, вместе с конкурентами, на других метриках, например, FIDclip, который подробно рассматривается в статье которую я упоминал выше “The Role of ImageNet Classes in Fréchet Inception Distance”.

Также сами авторы в качестве дальнейшей работы предлагают сделать 2 вещи: провести дистилляцию модели, найдя компромисс между качеством и производительностью, а также попробовать применить весь пайплайн на модель StyleGAN2, чтобы улучшить качество редактирования изображений.

***9. Попробуйте на основе результатов статьи предложить исследование, не проведенное к текущему моменту, или идею применение в индустриальных приложениях.***

StyleGAN-XL имеет большой потенциал по качеству генерации, но всё же возникают некоторые вопросы к количеству артефактов, я бы провёл дополнительное исследование по дообучению латентного пространства в более стабильное без артефактов, скорее всего, пожертвовав качеством и полнотой при генерации с метками, но подняв качество на более качественных метриках генерации, например, FIDclip. В частности я бы попробовал использовать PPL регуляризацию при дообучении или CLIP на некоторых наборах данных.

В индустрии StyleGAN-XL можно использовать как редактор человеческих лиц, так как он имеет лучшее качество генерации на FFHQ 1024 и при этом имеет хорошее латентное пространство.