



СОЗДАНИЕ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ ДЛЯ СИНТЕЗА РЕАЛИСТИЧНЫХ ИЗОБРАЖЕНИЙ: АРХИТЕКТУРА, МЕТОДЫ И ПЕРСПЕКТИВЫ

Семёнов Иван Алексеевич

старший преподаватель, Белорусский государственный университет
г. Минск, Беларусь.

Аннотация

Генеративные модели — одно из самых динамично развивающихся направлений в области искусственного интеллекта. Их способность создавать новые, ранее не существовавшие данные, нашла широкое применение в различных сферах: от дизайна и искусства до науки и медицины. Данная статья глубоко анализирует ключевые архитектуры и методы, используемые для синтеза реалистичных изображений, с особым акцентом на **генеративно-сопоставительные сети (GAN)** и **диффузионные модели**. Рассматриваются их принципы работы, эволюция, а также основные проблемы, связанные с обучением и оценкой качества. Подробно анализируются современные подходы к оптимизации генеративных моделей и их потенциал для будущих инноваций.

Ключевые слова: генеративные модели, GAN, диффузионные модели, синтез изображений, машинное обучение, искусственный интеллект, deep learning.

1. Введение: Фундаментальная задача генеративного моделирования

Генеративное моделирование — это раздел машинного обучения, посвященный обучению алгоритмов созданию новых данных, которые по своим характеристикам схожи с обучающей выборкой. Если задача дискриминативного моделирования (например, классификация изображений) сводится к предсказанию метки для заданных данных, то генеративное моделирование решает обратную задачу: создание новых экземпляров данных из исходного распределения. Эта способность открыла широкие перспективы для компьютерной графики, обработки изображений, создания контента для видеоигр и даже для генерации синтетических данных для обучения других моделей.

2. Генеративно-сопоставительные сети (GAN)

Генеративно-сопоставительные сети (GAN), предложенные Яном Гудфеллоу в 2014 году, стали революционным прорывом в области генеративного

моделирования. Их архитектура основана на "игре" между двумя нейронными сетями.

Архитектура GAN:

Генератор (Generator): Нейронная сеть, которая принимает на вход случайный "шум" (вектор латентного пространства) и преобразует его в изображение. Его цель — создать настолько реалистичное изображение, чтобы обмануть дискриминатор.

Дискриминатор (Discriminator): Нейронная сеть, которая принимает на вход изображение и пытается определить, является ли оно "настоящим" (из обучающей выборки) или "поддельным" (созданным генератором). Его цель — научиться отличать подделки.

Процесс обучения: Эти две сети обучаются одновременно в режиме состязания. Генератор постоянно улучшает свои способности к созданию реалистичных изображений, а дискриминатор — к их распознаванию. Этот процесс продолжается до тех пор, пока генератор не сможет создавать изображения, которые дискриминатор не в состоянии отличить от настоящих.

Проблемы GAN:

Режимный коллапс (Mode Collapse): Генератор может "застрять" на создании очень ограниченного набора изображений, которые гарантированно обманывают дискриминатор, игнорируя остальное разнообразие данных.

Нестабильность обучения: Процесс обучения GAN крайне чувствителен к гиперпараметрам и часто бывает нестабильным, что затрудняет получение высококачественных результатов.

3. Диффузионные модели: Новый подход к синтезу изображений

Диффузионные модели (Diffusion Models), такие как DALL-E 2, Midjourney и Stable Diffusion, стали новым стандартом в генерации изображений. Их подход кардинально отличается от GAN.

Принципы работы:

Прямой процесс (Forward Process): Модель постепенно добавляет случайный шум к реальному изображению, пока оно не превратится в чистый шум. Этот процесс полностью детерминирован.

Обратный процесс (Reverse Process): Модель обучается обращать этот процесс, шаг за шагом удаляя шум из зашумленного изображения, чтобы восстановить исходное. Этот процесс является генеративным.

Преимущества диффузионных моделей:

Высокое качество и разнообразие: Диффузионные модели известны своей способностью создавать невероятно детализированные и разнообразные изображения, избегая проблемы режимного коллапса, характерной для GAN.

Стабильность обучения: Обучение диффузионных моделей гораздо более стабильно и предсказуемо по сравнению с GAN.

4. Методы оптимизации и оценки

Для повышения эффективности и качества генеративных моделей используются различные методы.

Оценка качества:

Inception Score (IS): Оценивает качество и разнообразие сгенерированных изображений. Чем выше IS, тем лучше модель.

Fréchet Inception Distance (FID): Более точная метрика, которая сравнивает распределение сгенерированных изображений с распределением реальных. Чем ниже FID, тем выше качество.

Оптимизация:

Условные GAN (Conditional GAN): Позволяют управлять процессом генерации, например, генерировать изображение определенного класса (например, "кошка") или стиля.

Прогрессивно растущие GAN (Progressive GAN): Позволяют генерировать изображения высокого разрешения, начиная с маленьких и постепенно добавляя новые слои в сеть.

5. Перспективы и будущее генеративного моделирования

Синтез реалистичных изображений продолжает развиваться с невероятной скоростью.

Перспективы:

Создание видеоконтента: Генеративные модели уже используются для создания коротких видеороликов и анимации, а в будущем смогут генерировать полнометражные фильмы.

Медицина: Генерация синтетических медицинских изображений (например, рентгеновских снимков) для обучения диагностических моделей, что позволит решить проблему ограниченности данных.

Персонализация и дизайн: Создание уникальных аватаров, одежды и других объектов для виртуальных миров и метавселенных.

Вызовы:

Этические вопросы: Проблема поддельных изображений (deepfakes) ставит серьезные этические вопросы, связанные с дезинформацией и злоупотреблением технологией.

Вычислительные ресурсы: Обучение крупных генеративных моделей требует огромных вычислительных мощностей, что делает их недоступными для большинства исследователей и организаций.

Заключение

Генеративные модели, такие как GAN и диффузионные модели, совершили настоящий прорыв в области синтеза изображений. GAN заложили основы, предложив идею состязательного обучения, а диффузионные модели вывели качество и стабильность генерации на новый уровень. По мере развития технологий, эти модели будут играть все более важную роль в различных сферах, открывая невиданные ранее возможности для творчества и инноваций.

Литература

1. Goodfellow, I., et al. Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, **2014**, 27.
2. Dhariwal, P., & Nichol, A. Diffusion models beat GANs on image synthesis. *Advances in Neural Information Processing Systems*, **2021**, 34, 878-894.
3. Karras, T., et al. A style-based generator architecture for generative adversarial networks. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, **2019**, 4401-4410.
4. Ho, J., et al. Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, **2020**, 33, 6840-6851.
5. Radford, A., et al. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, **2015**.
6. Heusel, M., et al. GANs trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in Neural Information Processing Systems*, **2017**, 30.
7. Karras, T., et al. Progressive growing of GANs for improved quality, stability, and variation. *International Conference on Learning Representations*, **2018**.
8. Sonderby, S. K., et al. Amortised inference in variational autoencoders with importance sampling. *arXiv preprint arXiv:1608.06830*, **2016**.
9. Ramesh, A., et al. Hierarchical text-conditional image generation with DALL-E. *arXiv preprint arXiv:2102.10091*, **2021**.
10. Nichol, A. Q., et al. Improved denoising diffusion probabilistic models. *arXiv preprint arXiv:2102.09672*, **2021**.