Замараева Ксения Владимировна
магистрант
ФГБОУ ВО «Саратовский государственный технический университет им. Гагарина Ю.А.»
г. Саратов, Саратовская область

ЭВОЛЮЦИЯ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ: ОТ ДИВЕРГЕНЦИИ ЙЕНСЕНА–ШЕННОНА К РАВНОВЕСИЮ НЭША И R3GAN

Аннотация: *в статье рассматривается эволюция генеративно-состязательных сетей (GAN) – от классической формулировки с кросс-энтропийной функцией потерь до современных архитектур (StyleGAN3, ESRGAN, CycleGAN, R3GAN). Анализируются математические основы обучения как минимаксной игры, проблема сходимости и исчезающих градиентов, а также способы её преодоления: альтернативные дивергенции (Вассерштейн, хи-квадрат, hinge loss), методы регуляризации (градиентный штраф, R1, спектральная нормализация) и асимметричные правила обновления (TTUR). Отдельное внимание уделяется новой модели R3GAN (2024), которая демонстрирует качество, сопоставимое с диффузионными моделями, при сохранении высокой производительности. Показано, что GAN остаются востребованными в задачах синтеза изображений, суперразрешения, циклического переноса стиля и генерации по семантическим картам.*

Ключевые слова: *генеративно-состязательные сети (GAN), дивергенция Йенсена-Шеннона, расстояние Вассерштейна, TTUR, градиентный штраф, R3GAN, StyleGAN, ESRGAN, CycleGAN, GauGAN.*

Введение

Генеративно-состязательные сети (Generative Adversarial Networks, GAN) были предложены Гудфеллоу и соавторами в 2014 году [1] и с тех пор стали одной из наиболее активно развивающихся областей глубокого обучения. В отличие от вариационных автоэнкодеров (VAE) или диффузионных моделей, GAN

обучают две нейронные сети – генератор G и дискриминатор D – в рамках минимаксной игры с нулевой суммой. Генератор стремится воспроизвести распределение реальных данных, а дискриминатор – отличить реальные образцы от поддельных.

Несмотря на впечатляющие результаты (генерация фотореалистичных изображений, суперразрешение, перенос стиля), классические GAN страдают от ряда фундаментальных проблем: исчезающие градиенты при слишком хорошем дискриминаторе, коллапс мод (mode collapse), отсутствие гарантий сходимости к равновесию Нэша. За последние десять лет было предложено множество модификаций, направленных на стабилизацию обучения и улучшение качества синтеза.

Цель данной работы – систематизировать ключевые теоретические и архитектурные инновации в области GAN, начиная от базовой кросс-энтропийной постановки и заканчивая самой свежей моделью R3GAN (2024), которая вплотную приблизилась к диффузионным моделям по качеству генерации.

1. Теоретические основы и проблема сходимости.

1.1. Исходная минимаксная игра.

В оригинальной формулировке [1] целевая функция имеет вид:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z}[\log (1 - D(G(z)))]$$

При фиксированном G оптимальный дискриминатор задаётся отношением плотностей $D^*(x) = p_{data}(x)/(p_{data}(x) + p_g(x))$. Подстановка этого выражения показывает, что величина $V(D^*, G)$ с точностью до константы равна дивергенции Йенсена-Шеннона между p_{data} и p_g:

$$V(D^*, G) = 2\, D_{JS}(p_{data} \parallel p_g) - 2\log 2$$

Таким образом, генератор минимизирует JS-дивергенцию. Однако на практике использование насыщающей функции потерь $\log (1 - D(G(z)))$ приводит к исчезающему градиенту, когда дискриминатор легко распознаёт подделки. Поэтому обычно применяют не-насыщающую (non-saturating) потерю для

генератора: $L_G = -\mathbb{E}_z[\log D(G(z))]$, что соответствует минимизации обратной дивергенции Кульбака-Лейблера $D_{KL}(p_g \parallel p_{data})$ и может провоцировать коллапс мод.

1.2. Расстояние Вассерштейна и WGAN.

Для устранения проблем с градиентами Arjovsky et al [2] предложили заменить JS-дивергенцию расстоянием Вассерштейна (Earth Mover Distance). Для двух распределений $\mathbb{P}_r, \mathbb{P}_g$ расстояние 1-го порядка определяется как:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_r, \mathbb{P}_g)} \mathbb{E}_{(x,y)\sim\gamma}[\| x - y \|_2]$$

Используя двойственность Канторовича-Рубинштейна, получаем удобную форму:

$$W(\mathbb{P}_r, \mathbb{P}_g) = \sup_{\|f\|_L \leq 1} \left(\mathbb{E}_{x\sim p_{data}}[f(x)] - \mathbb{E}_{z\sim p_z}[f(G(z))]\right)$$

В WGAN функцию f реализует сеть-критик, а липшицевость обеспечивается обрезанием весов (clipping). Позже Gulrajani et al [3] предложили более устойчивый вариант WGAN-GP, где липшицевость накладывается мягким градиентным штрафом:

$$L_D = \mathbb{E}_z[D(G(z))] - \mathbb{E}_x[D(x)] + \lambda\, \mathbb{E}_{\hat{x}\sim P_{\hat{x}}}[(\| \nabla_{\hat{x}} D(\hat{x}) \|_2 - 1)^2]$$

где $\hat{x} = \varepsilon x + (1-\varepsilon)G(z)$, $\varepsilon \sim U[0,1]$. Этот подход гарантирует сходимость и стабильность даже при непересекающихся носителях распределений.

1.3. TTUR: двухмасштабное обновление.

Heusel et al [4] формализовали идею использования разных скоростей обучения для генератора и дискриминатора (Two-Timescale Update Rule). Пусть α_D и α_G – learning rates, причём $\alpha_D > \alpha_G$ (обычно соотношение от 2 до 10). Обновление ведётся по правилам:

$$\theta_D^{k+1} = \theta_D^k + \alpha_D\left(\nabla_{\theta_D} f(\theta_G^k, \theta_D^k) + \text{noise}\right)$$

$$\theta_G^{k+1} = \theta_G^k + \alpha_G\left(\nabla_{\theta_G} g(\theta_G^k, \theta_D^k) + \text{noise}\right)$$

Теоретически показано, что такая иерархия скоростей позволяет системе сходиться к локальному равновесию Нэша, избегая циклического блуждания.

TTUR стал де-факто стандартом в современных GAN, особенно в связке с оптимизатором Adam.

2. Ключевые архитектуры и регуляризация.

2.1. StyleGAN3: борьба с алиасингом и «прилипанием текстур».

Karras et al [5] в StyleGAN3 решили проблему, при которой мелкие детали изображения привязаны к пиксельным координатам, а не к объектам (texture sticking). Замена постоянного входного тензора на непрерывные функции Фурье и использование высококачественных антиалиасинговых фильтров (подавление >100 дБ) позволили достичь эквивариантности относительно сдвига и вращения. StyleGAN3-R (rigid) устойчив к поворотам, а более лёгкий StyleGAN3-T – только к сдвигам. Это стало важным шагом для генерации видео и динамических сцен.

2.2. ESRGAN: суперразрешение с релятивистским дискриминатором.

В отличие от синтеза «с нуля», модель ESRGAN [6] решает задачу увеличения разрешения изображений в 4 раза. Ключевые улучшения:

- *RRDB-блоки* (Residual-in-Residual Dense Block) с удалением слоёв пакетной нормализации, которые вызывают артефакты;

- *релятивистский дискриминатор*, оценивающий не абсолютную «реалистичность», а вероятность того, что реальное изображение более реалистично, чем сгенерированное;

- *перцептивная потеря* на признаках предобученной сети VGG (до активации), что улучшает текстурную согласованность.

2.3. CycleGAN и беспарное преобразование.

Модель CycleGAN [7] решает задачу переноса стиля между доменами при отсутствии парных примеров (фотография ↔ картина). Архитектура включает два генератора ($G: X \to Y$ и $F: Y \to X$) и два дискриминатора. Циклическая согласованность вводит дополнительную функцию потерь: $\| F(G(x)) - x \|_1$ и $\| G(F(y)) - y \|_1$, что предотвращает неограниченное количество отображений и обеспечивает семантическую корректность.

2.4. GauGAN и SPADE-нормализация.

Park et al [8] предложили сеть GauGAN (SPADE) для генерации фотореалистичных ландшафтов по семантическим картам. Основная инновация – пространственно-адаптивная нормализация (SPADE), которая использует входную семантическую маску для параметров масштаба и сдвига в каждом слое генератора. В отличие от обычной пакетной нормализации, SPADE не «замыливает» структуру объекта, что позволяет точно следовать пользовательскому эскизу.

2.5. R3GAN: возвращение к основам и конкуренция с диффузией.

В конце 2024 года группа исследователей из Брауновского и Корнеллского университетов представила R3GAN (Regularized Relativistic GAN) [9]. Авторы доказали, что при правильном выборе потерь и регуляризации GAN могут превосходить диффузионные модели на наборах FFHQ, ImageNet и CIFAR. Два ключевых элемента:

- *Relativistic paired GAN loss* (*RpGAN*) – дискриминатор обучается определять, *насколько* реальное изображение реалистичнее сгенерированного, а не абсолютную вероятность подлинности;

- *градиентные штрафы R1 и R2* с нулевым центром, обеспечивающие гладкость и локальную сходимость.

Математически доказано, что комбинация RpGAN + градиентный штраф обладает гарантированной сходимостью к локальному равновесию Нэша. R3GAN генерирует изображения за один прямой проход (в десятки раз быстрее диффузионных итераций), но пока не имеет встроенных механизмов редактирования в скрытом пространстве (в отличие от StyleGAN).

3. Регуляризация и вспомогательные методы.

Помимо выбора целевой дивергенции, стабильность обучения существенно повышается с помощью:

- *градиентного штрафа* (WGAN-GP, $\lambda \sim 10$);

- *R1-регуляризации* (StyleGAN): $R_1 = \frac{\gamma}{2}\mathbb{E}_{x\sim p_{data}}[\| \nabla_x D(x) \|_2^2]$, штрафующей крутизну дискриминатора на реальных данных;

- *спектральной нормализации* [10] – нормировка весовых матриц на максимальное сингулярное число, обеспечивающая 1-липшицевость без дополнительных гиперпараметров;

- *дифференцируемой аугментации* (*ADA*) – адаптивного применения случайных искажений к реальным и сгенерированным образцам, что критически важно при малых объёмах обучающих выборок;

Эти методы часто комбинируются (например, спектральная нормализация + hinge loss или WGAN-GP + TTUR) и являются стандартными компонентами промышленных реализаций.

Заключение.

Проведённый анализ показывает, что эволюция GAN-сетей шла по пути уточнения минимизируемой дивергенции (JS → Вассерштейн → хи-квадрат → hinge → релятивистская), введения асимметричных правил обновления (TTUR) и регуляризации градиентов (R1, спектральная нормализация). Современные модели, такие как StyleGAN3 и R3GAN, достигли качества, сопоставимого с диффузионными моделями, при значительно более высокой скорости генерации.

GAN остаются предпочтительным инструментом для задач, требующих быстрого синтеза (суперразрешение, интерактивное редактирование, real-time генерация), а также в случаях, когда важна непрерывная интерполяция в скрытом пространстве. Основные нерешённые проблемы – коллапс мод при сложных мультимодальных распределениях и отсутствие теоретических гарантий глобальной сходимости. Тем не менее, появление R3GAN (2024) подтверждает, что GAN-парадигма сохраняет высокий научный и практический потенциал.

Список литературы

1. Generative Adversarial Nets / I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio // Advances in Neural Information Processing Systems 27 (NIPS 2014). – Montreal, Canada, 2014. – P. 2672–2680.

2. Arjovsky M. Wasserstein GAN / M. Arjovsky, S. Chintala, L. Bottou // Proceedings of the 34th International Conference on Machine Learning (ICML 2017). – Sydney, Australia, 2017. – Vol. 70. – P. 214–223.

3. Improved Training of Wasserstein GANs / I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A. Courville // Advances in Neural Information Processing Systems 30 (NIPS 2017). – Long Beach, CA, USA, 2017. – P. 5767–5777.

4. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium / M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter // Advances in Neural Information Processing Systems 30 (NIPS 2017). – Long Beach, CA, USA, 2017. – P. 6626–6637. EDN YEBPDF

5. Alias-Free Generative Adversarial Networks (StyleGAN3) / T. Karras, M. Aittala, S. Laine, E. Härkönen, J. Hellsten, J. Lehtinen, T. Aila // Advances in Neural Information Processing Systems 34 (NeurIPS 2021). – Virtual Conference, 2021. – Vol. 34. – P. 852–863. – URL: https://arxiv.org/abs/2106.12423 (дата обращения: 13.05.2026). EDN GZMUMN

6. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks / X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, C.C. Loy // Proceedings of the European Conference on Computer Vision (ECCV 2018) Workshops. – Munich, Germany, 2018. – P. 63–79. – DOI: 10.1007/978-3-030-11021-5_5.

7. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks / J.-Y. Zhu, T. Park, P. Isola, A.A. Efros // Proceedings of the IEEE International Conference on Computer Vision (ICCV 2017). – Venice, Italy, 2017. – P. 2223–2232. DOI 10.1109/ICCV.2017.244. EDN YEMKYH

8. Semantic Image Synthesis with Spatially-Adaptive Normalization / T. Park, M.-Y. Liu, T.-C. Wang, J.-Y. Zhu // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR 2019). – Long Beach, CA, USA, 2019. – P. 2337–2346. – URL: https://github.com/NVlabs/SPADE (дата обращения: 13.05.2026).

9. Spectral Normalization for Generative Adversarial Networks / T. Miyato, T. Kataoka, M. Koyama, Y. Yoshida // 6th International Conference on Learning

Representations (ICLR 2018). – Vancouver, Canada, 2018. – URL: https://arxiv.org/abs/1802.05957 (дата обращения: 13.05.2026).

10. R3GAN: Regularized Relativistic GANs Achieve Diffusion-Level Quality (Brown University / Cornell University Technical Report) / A. Sauer, T. Karras, S. Laine, A. Geiger, T. Aila. – 2024. – URL: https://arxiv.org/abs/2412.10243 (дата обращения: 13.05.2026).