

УДК 51-37

ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫЕ НЕЙРОННЫЕ СЕТИ: ИСТОРИЯ РАЗВИТИЯ И ОПИСАНИЕ ГЕНЕРАТИВНЫХ МОДЕЛЕЙ

Нагорный Максим Алексеевич,

Уральский государственный экономический университет, студент

Электронная почта: maxym.nagorny.61@gmail.com

Россия, Екатеринбург

Аннотация

В статье рассмотрены ранние проблемы генеративных нейронных сетей и их дальнейшее развитие, опирающееся на архитектуру генеративно-сопоставительных нейронных сетей. Описаны различные типы генеративно-сопоставительных нейронных сетей с сопоставлением к применяемым областям. Также приведены конкретные примеры, демонстрирующие качество развившихся генеративных моделей.

Ключевые слова: GAN, архитектура нейронных сетей, генеративные нейронные сети, генерация изображений, генерация музыки, генерация текстов, глубокое машинное обучение.

GENERATIVE-ADVERSARIAL NEURAL NETWORKS: HISTORY OF DEVELOPMENT AND DESCRIPTION OF GENERATIVE MODELS

Nagorny Maxim Alekseevich,

Ural State University of Economics, student

Email: maxym.nagorny.61@gmail.com

Russia, Yekaterinburg

ABSTRACT

This article reviews the early problems of generative neural networks and their further development based on the generative-adversarial neural network architecture. Different types of generative-adversarial neural networks are described with a comparison to the fields of application. Specific examples are also provided to demonstrate the quality of existing generative models.

Keywords: GAN, neural network architecture, generative neural networks, image generation, music generation, text generation, deep machine learning.

Перспективой применения глубокого обучения является её использование для открытия новых сложных иерархических моделей, которые позволяют распределять вероятности принадлежности тому или иному классу у комплексных видов данных и которые встречаются в приложениях с искусственным интеллектом. Речь идёт про такие данные, как изображения, звуковые волны с содержанием речи и символы в корпусах естественного языка. Около десяти лет назад наиболее яркие успехи в глубоком обучении были связаны в основном с дискриминационными моделями, которые сопоставляют многомерные данные, воспринимаемые органами чувств человека, с соответствующими метками классов. Эти поразительные успехи в основном основаны на алгоритмах обратного распространения (backpropagation) и исключения (dropout).

Глубокие генеративные модели оказывали меньше влияния из-за сложности замены многих трудновыполнимых вероятностных вычислений, возникающих при оценке максимального соответствия новых данных с исходными, более простыми, а также из-за сложности использования имеющихся алгоритмов в контексте генерации данных. В 2014 году была предложена процедура оценки генеративной модели, которая позволила бы обойти эти трудности, о ней и пойдёт речь далее [3].

Генеративно-сопоставительная нейронная сеть или GAN (Generative Adversarial Net), — это генеративная модель, основанная на глубоком обучении. В более общем смысле GAN — это архитектура обучения генеративной модели [4].

Впервые эта архитектура была описана в 2014 году в работе Яна Гудфеллоу и других авторов под названием «Generative Adversarial Networks» [3]. В ней предлагается структура из двух моделей, этапы работы которой можно описать следующим образом:

1. Инициализация: создаются две нейронные сети: генератор и дискриминатор. Задача генератора — создавать новые данные, такие как изображения или текст, которые очень похожи на реальные данные. Дискриминатор же выступает в роли критика, пытаясь отличить реальные данные (из обучающего набора данных) от данных, сгенерированных генератором.

2. Первый ход генератора: он принимает на вход вектор случайного шума (набор случайных чисел). Он служит отправной точкой для процесса создания новых данных. Используя свои внутренние слои и выученные паттерны, генератор преобразует вектор шума в новый образец данных, например: сгенерированное изображение.

3. Подключение к процессу дискриминатора, он получает два вида входных данных:
- Реальные образцы из обучающего набора данных;
- Образцы, сгенерированные генератором на предыдущем шаге.

Его задача — проанализировать каждый входной сигнал и определить, реальные это данные или сгенерированные. Он выдает оценку вероятности от 0 до 1. Оценка 1 означает, что данные, скорее всего, реальные, а 0 — поддельные.

4. Запуск процесса обучения, теперь в дело вступает сопоставительная часть: если дискриминатор правильно определяет реальные данные как настоящие (оценка близка к 1), а сгенерированные как поддельные (оценка близка к 0), то и обе сети получают небольшое вознаграждение. Это происходит из-за того, что обе сети хорошо выполняют свою работу.

Поэтому очень важно постоянно их совершенствовать. Если дискриминатор будет постоянно определять все правильно, он не сможет многому научиться. Цель при обучении состоит в том, чтобы генератор смог убедить своего соперника в реальности сгенерированных данных.

5. Совершенствование генератора: когда дискриминатор ошибочно называет новые данные реальными (оценка близка к 1), это признак того, что генератор на правильном пути. В этом случае дискриминатор получает штраф за то, что его обманули, а его

конкурент в свою очередь наоборот получает вознаграждение. Такая связь помогает генератору улучшить качество новых данных, чтобы они больше соответствовали реальным.

6. Адаптация дискриминатора: эта связь работает и в обратном направлении: если дискриминатор правильно определит поддельные данные (оценка близка к 0), то таким образом он показывает качество распознавания данных, а генератор не получает награды. Эта постоянная дуэль между двумя сетями со временем их совершенствует [6].

Общую структуру процесса обучения генеративно-сопоставительной нейронной сети можно увидеть на рисунке 1.

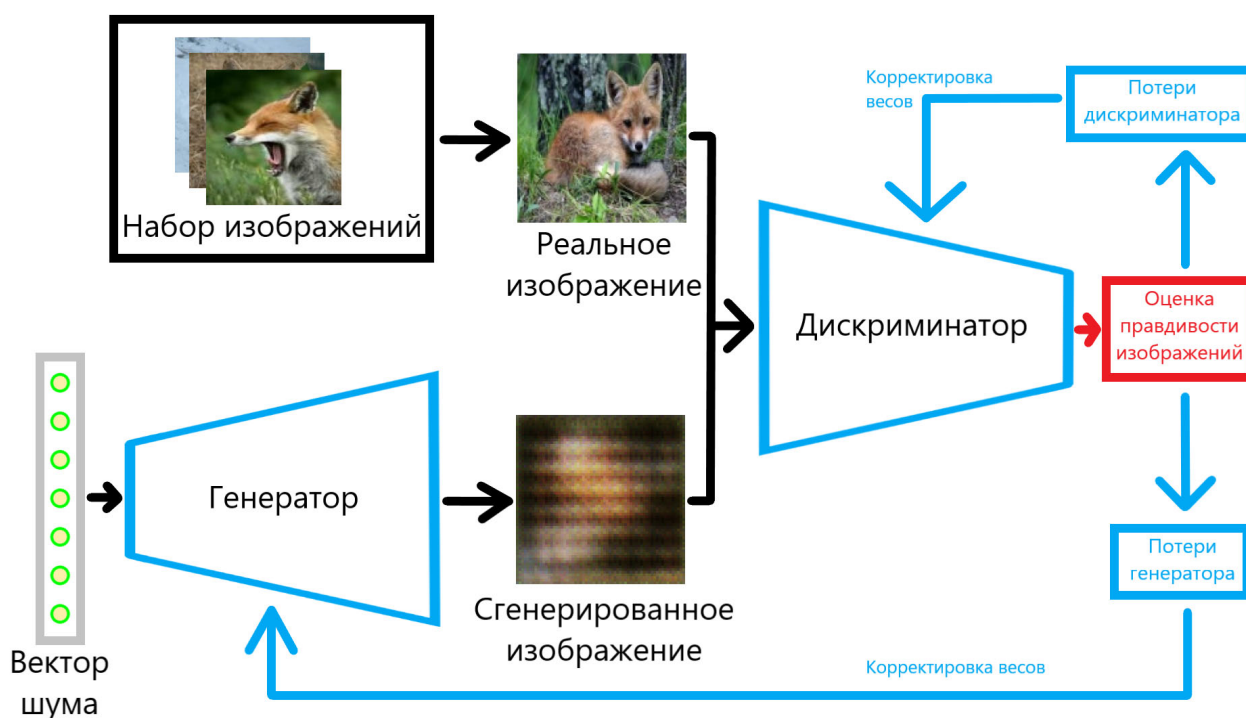


Рисунок 1 – процесс обучения GAN.1

На основе архитектуры генеративно-сопоставительных нейронных сетей позже появился другой стандартизированный подход под названием «Глубокая свёрточная генеративно-сопоставительная нейронная сеть» (Deep Convolutional Generative Adversarial Network или DCGAN), который привел к созданию более стабильных моделей. Подход был описан Алемом Рэдфордом и другими авторами в работе 2015 года под названием «Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks». В ней предлагается сделать следующие изменения в классической структуре GAN:

- Заменить все пулинг-слои разреженными свёрточными слоями (в дискриминаторе) и разреженными свёрточными слоями свертки с дробными шагами (в генераторе).
 - Использовать пакетную нормализацию как в генераторе, так и в дискриминаторе.
 - Удалить полносвязные скрытые слои для более глубоких архитектур.
 - Использовать функцию активации ReLU в генераторе для всех слоев, кроме выходного, там использовать Tanh.
 - Использовать функцию активации LeakyReLU в дискриминаторе для всех слоев [1].
- Помимо этих двух представлений, GAN также имеет другие типы архитектур:
- Условная GAN (CGAN). Применяя метки классов, этот тип позволяет обеспечить сеть новой и классифицированной информацией. В процессе обучения GAN сеть получает

¹ Рисунок 1 составлен автором

изображения с их реальными метками, такими как «роза», «подсолнух» или «тюльпан», чтобы помочь ей научиться различать их.

- Циклически-согласованная GAN (CycleGAN). Это наиболее распространенная архитектура GAN, которая обычно используется для обучения преобразованию изображений в другие стили. Например: модель можно научить изменять изображение с зимнего на летнее или с панды на красную панду. Также модель может визуально изменять возраст человеческих лиц.

- StyleGAN. Исследователи из Nvidia выпустили StyleGAN в декабре 2018 года и предложили значительные улучшения оригинальной архитектуры генератора моделей [7]. Эта архитектура может создавать фотореалистичные, высококачественные фотографии лиц, но пользователи могут модифицировать модель, чтобы изменить внешний вид получаемых изображений.

- GAN для высокого разрешения (SRGAN). Благодаря этой архитектуре изображение с низким разрешением может быть преобразовано в более детализированное. Модели с этой архитектурой увеличивают разрешение изображения, заполняя размытые участки [5].

Появление классической архитектуры генеративно-состязательной нейронной сети привело к большому развитию генеративных моделей и архитектур, которые уже достигли высокого качества в генерации изображений, текстов и музыки, из наиболее ярких примеров можно выделить Stable Diffusion, ChatGPT и Suno AI. Конкретно эти примеры используют относительно гораздо более сложную архитектуру под названием «Трансформер», которая впервые была предложена исследователями в рамках проекта Google Brains в 2017 году, спустя 3 года после появления классического GAN [2]. Эту архитектуру также можно усложнять и улучшать качество генерируемых данных, и сейчас это особенно актуально, ввиду возросшего интереса к подобным инструментам.

Автор данной статьи выражает благодарность доктору экономических наук, доценту, заведующему кафедрой бизнес-информатики Уральского государственного экономического университета Назарову Дмитрию Михайловичу за оценку качества написанного материала и составленного изображения структуры обучения генеративно-состязательной нейронной сети в роли научного руководителя.

Список литературы:

1. Alec Radford // Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks - [2016]: науч. ст. - URL: <https://arxiv.org/abs/1511.06434/> / (дата обращения: 21.03.2025).
2. Ashish Vaswani // Attention Is All You Need - [2017]: науч. ст. - URL: <https://arxiv.org/abs/1706.03762/> / (дата обращения: 22.03.2025).
3. Ian J. Goodfellow // Generative Adversarial Networks - [2014]: науч. ст. - URL: <https://arxiv.org/abs/1406.2661/> / (дата обращения: 21.03.2025).
4. Jason Brownlee // A Gentle Introduction to Generative Adversarial Networks (GANs) - [2019]: сайт. - URL: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> / (дата обращения: 21.03.2025).
5. Kinza Yasar // generative adversarial network (GAN) - [2023]: сайт. - URL: <https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN/> / (дата обращения: 21.03.2025).
6. Rahul Roy // Generative Adversarial Network (GAN) - [2024]: сайт. - URL: <https://www.geeksforgeeks.org/generative-adversarial-network-gan/> / (дата обращения: 21.03.2025).

7. Tero Karras // A Style-Based Generator Architecture for Generative Adversarial Networks - [2018]: науч. ст. - URL: <https://arxiv.org/abs/1812.04948> / (дата обращения: 24.03.2025).

References:

1. Alec Radford // Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks - [2016]: research article - URL: <https://arxiv.org/abs/1511.06434/> / (accessed: 21.03.2025).
2. Ashish Vaswani // Attention Is All You Need - [2017]: research article - URL: <https://arxiv.org/abs/1706.03762/> / (accessed: 22.03.2025).
3. Ian J. Goodfellow // Generative Adversarial Networks - [2014]: research article - URL: <https://arxiv.org/abs/1406.2661/> / (accessed: 21.03.2025).
4. Jason Brownlee // A Gentle Introduction to Generative Adversarial Networks (GANs) - [2019]: website. - URL: <https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/> / (access date: 03/21/2025).
5. Kinza Yasar // generative adversarial network (GAN) - [2023]: website. - URL: <https://www.techtarget.com/searchenterpriseai/definition/generative-adversarial-network-GAN/> / (access date: 03/21/2025).
6. Rahul Roy // Generative Adversarial Network (GAN) - [2024]: website. - URL: <https://www.geeksforgeeks.org/generative-adversarial-network-gan/> / (date of access: 21.03.2025).
7. Tero Karras // A Style-Based Generator Architecture for Generative Adversarial Networks - [2018]: scientific. article - URL: <https://arxiv.org/abs/1812.04948> / (date of access: 24.03.2025).