DOI 10.37882/2223-2966.2024.02.13

НЕЙРОСЕТЬ STABLE DIFFUSION: ОСОБЕННОСТИ АРХИТЕКТУРЫ

STABLE DIFFUSION NEURAL NETWORK: ARCHITECTURE FEATURES

N. Verezubova
A. Chekulaev
I. Verezubova

Summary. This article discusses the Stable Diffusion neural network and the features of its architecture. Stable Diffusion is a generative — adversarial model developed in 2022 and has been actively developing since then. It is used to create and modify images using a generator and discriminator that work together to improve image quality.

Keywords: neural network, Stable Diffusion, generative-adversarial models, machine learning, U-Net, ResNet, CLIP model, on-demand image generation scheme.

Верезубова Наталья Афанасьевна
Кандидат экономических наук, доцент, Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
nverez@mail.ru
Чекулаев Артур Анатольевич
Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина
Верезубова Ирина Николаевна
Московская государственная академия ветеринарной медицины и биотехнологии имени К.И. Скрябина

Аннотация. В данной статье рассматривается нейросеть Stable Diffusion и особенности ее архитектуры. Stable Diffusion — это генеративно-состязательная модель, разработанная в 2022 году и активно развивающаяся с тех пор. Она используется для создания и изменения изображений с помощью генератора и дискриминатора, которые работают вместе для улучшения качества изображений.

Ключевые слова: нейросеть, Stable Diffusion, генеративно-состязательные модели, машинное обучение, U-Net, ResNet, модель CLIP, схема генерации картинок по запросу.

Нейросети являются одним из ключевых инструментов в области искусственного интеллекта. Они представляют собой математическую модель, которая имитирует работу человеческого мозга. Нейросеть состоит из множества связанных между собой элементов, называемых нейронами, которые обрабатывают информацию.

Архитектура нейронных сетей описывает структуру и организацию нейронной сети, включая количество слоев, количество нейронов в каждом слое, функции активации, методы оптимизации и другие параметры, которые определяют, как сеть будет обрабатывать входные данные и выдавать результаты.

Существуют различные типы нейросетей, каждый из которых предназначен для решения определенной задачи. Например, сверточные нейронные сети используются для анализа медицинских изображений, рекуррентные нейросети применяются для анализа временных рядов в экономике, генеративно-состязательные нейросети используются для создания новых дизайнов в области моды или визуальных эффектов в киноиндустрии [1, 2].

Генеративно-состязательные нейронные сети (GAN) являются одним из самых перспективных направлений в области искусственного интеллекта и машинного обучения. Нейронные сети данного типа используются для создания новых данных и могут применяться в различных областях, например, таких как обработка изображений. Они состоят из двух частей: генератора и дискриминатора. Генератор создает новые данные, а дискриминатор пытается определить, является ли данное ему изображение настоящим или сгенерированным. Обучение происходит в процессе соревнования между генератором и дискриминатором. Такие нейросети могут использоваться для улучшения качества изображений, создания новых изображений из существующих, а также для определения трендов в данных. Однако они могут требовать большого количества времени и вычислительных ресурсов для обучения, и по запросу создают данные, которые не соответствуют реальным [3, с. 1–8]. Схема работы генеративно-состязательные нейронные сети представлена на рисунке 1.

Одним из представителей генеративно-состязательной сети является нейросеть Stable Diffusion.

Stable Diffusion — это одна из самых популярных и передовых нейросетей в сфере искусственного интеллекта и машинного обучения. Она была разработана в 2022 году компанией Stability.ai. Одним из основных преимуществ Stable Diffusion перед другими нейросетями является ее способность создавать изображения с высоким разрешением и детализацией. Это достигается благодаря использованию алгоритмов глубокого обучения и современных технологий, таких как U-Net и ResNet.

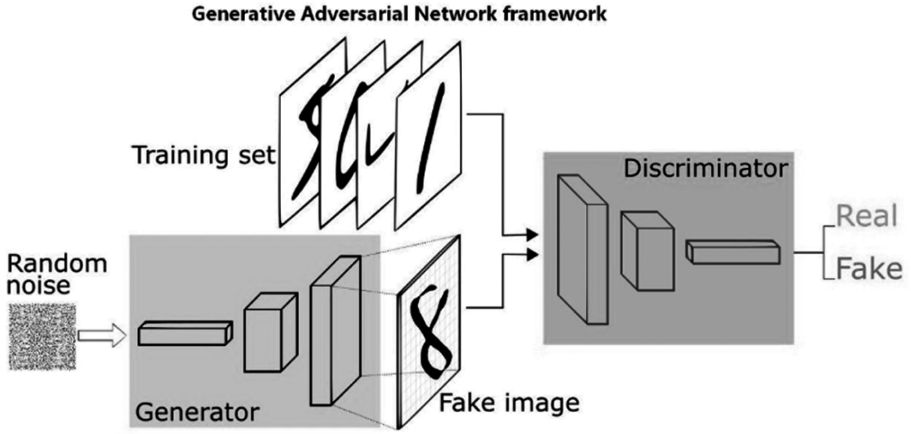

Рис. 1. Схема работы GAN [1]

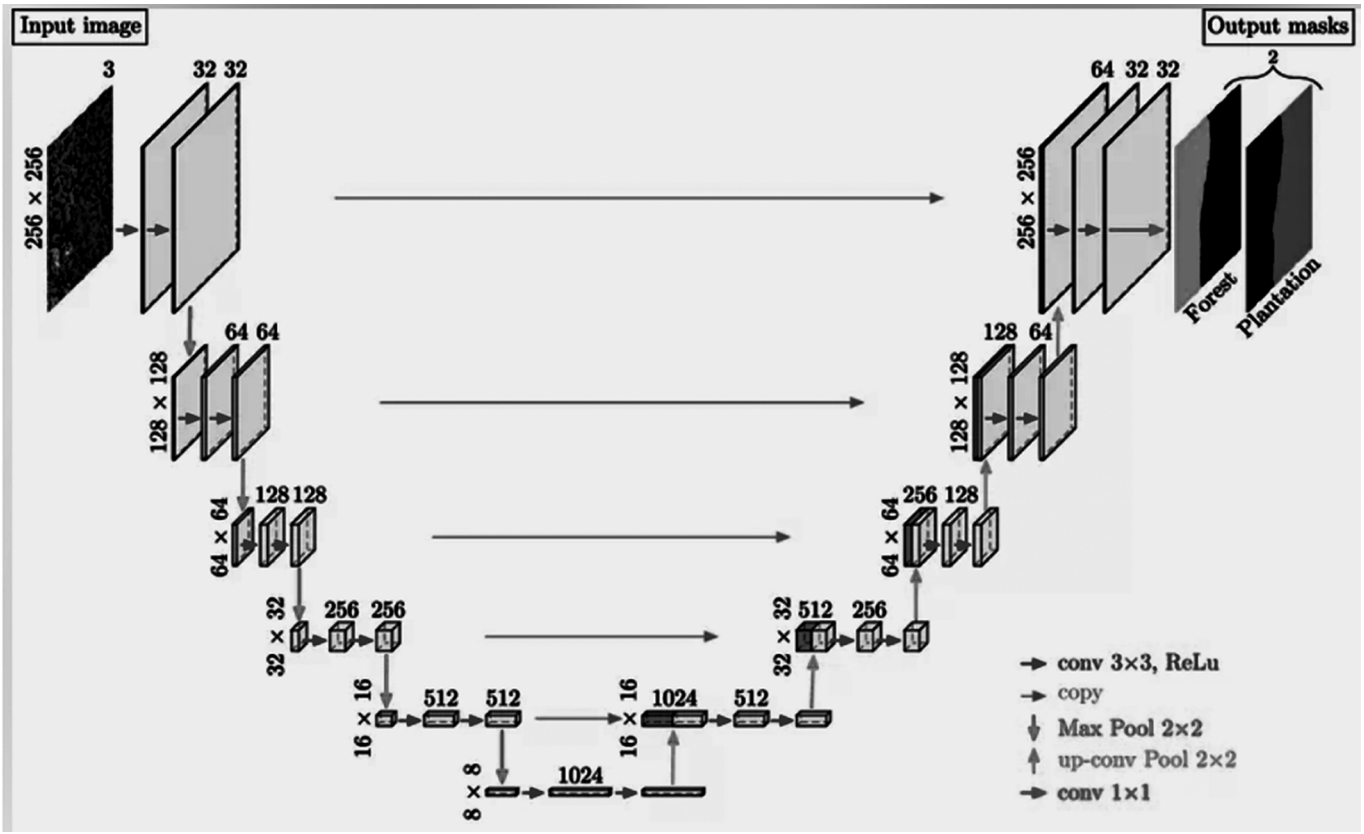

Рис. 2. Схема работы U-Net [5]

U-Net представляет собой сверточную нейронную сеть, которая была создана в 2015 году для сегментации биомедицинских изображений. Её Архитектура — это полносвязная свёрточная сеть, модифицированная так, чтобы она могла работать с меньшим количеством примеров (обучающих образов) и делала более точную сегментацию [4].

ResNet (сокращение от Residual Network) — это архитектура глубокой сверточной нейронной сети разработана для решения проблемы исчезающих градиентов, которая может возникнуть при обучении глубоких нейронных сетей. Идея ResNet заключается в добавлении коротких связей (также известных как пропускные связи или остаточные связи), которые обходят некоторые слои сети. Благодаря этому сеть может научиться повторно использовать характеристики, полученные в ранних слоях сети, в последующих слоях, даже если эти характеристики очень малы. Это помогает предотвратить то, что градиенты становятся слишком маленькими и исчезают совсем [6].

Обучение Stable Diffusion происходит на больших объемах данных, включающих в себя различные изображения и описания. В процессе обучения нейросеть учится сопоставлять текстовое описание с изображением

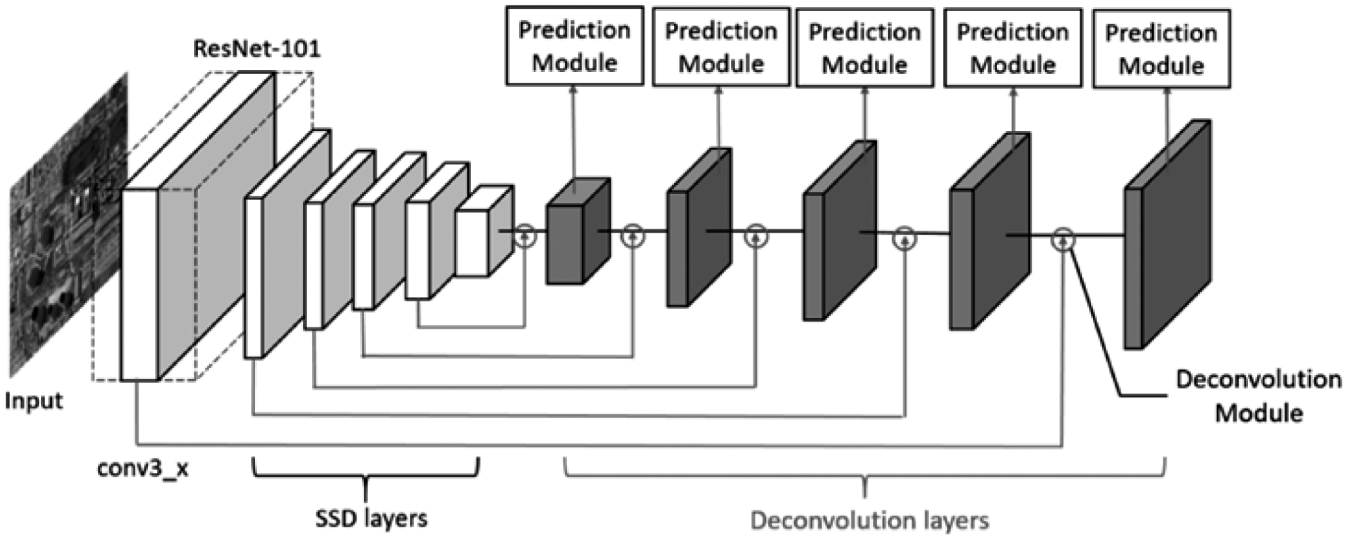

Рис. 3. Схема работы ResNet [7]

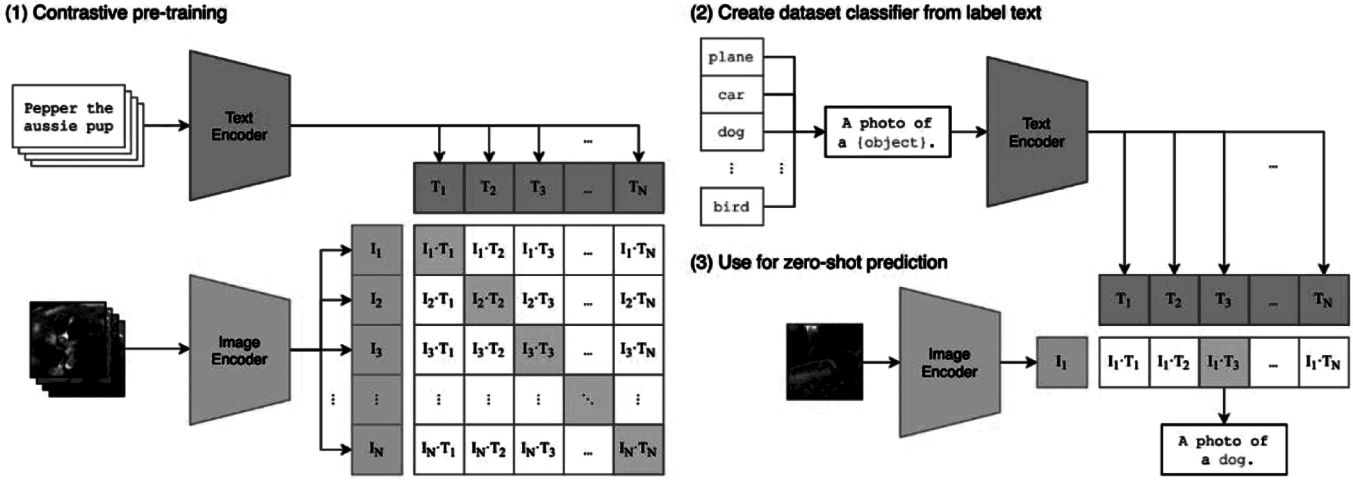

Рис. 4. Схема работы CLIP [9]

и создавать новые изображения на основе полученных знаний.

Несмотря на все преимущества Stable Diffusion имеет некоторые ограничения. Например, она может быть довольно медленной при создании изображений и требует значительных вычислительных ресурсов. Кроме того, качество изображений может зависеть от точности описания и правильности настройки нейросети.

Анализ архитектуры Stable Diffusion следует начать с исследования ее ключевых компонентов. Первоначально рассмотрим кодировщик исходных данных, который трансформирует текстовый запрос в векторный признак. Далее нужно перейти к генератору интермедиатных изображений, который генерирует промежуточные изображения на основании исходного текста. Репозиторий данной нейросети можно найти на Git Hub.

Стоит отметить, что у каждого из этих компонентов есть свои настраиваемые параметры, которые влияют на итоговый результат. К примеру, можно менять размер скрытого слоя, число слоев или использовать разные функции активации.

Проведем анализ кодировщика исходных данных и генератора промежуточных изображений. Скрипты написаны на языке Python и использует библиотеку Stable Diffusion для работы с нейросетями.

Код скрипта начинается с импорта необходимых библиотек и определения функций, которые будут использоваться для генерации изображений. Затем идет основная часть кода, которая обрабатывает входной текст и генерирует изображения.

Модель CLIP (Contrastive Language-Image Pre-training) — это тип нейросетевого алгоритма, разработанный компанией OpenAI. Он способен анализировать взаимосвязь между визуальными образами и текстовыми описаниями на естественном языке.

CLIP является техническим нововведением применения двунаправленного обучения трансформеров в языковом моделировании. Этот подход отличается от предыдущих моделей, которые рассматривали текстовую последовательность либо только слева направо, либо сочетали обучение слева направо и справа налево. Языковая модель с двунаправленным обучением способна достичь более глубокого понимания языкового контекста и потока, чем однонаправленные языковые модели

Созданная для обучения нейросетей распознаванию и связыванию широкого спектра изображений с их описаниями, модель CLIP прошла предварительное обучение на большом наборе данных с использованием алгоритма контрастного обучения [8].

Далее код использует нейросеть Stable Diffusion для генерации изображения на основе полученного вектора текста.

После генерации изображения код оптимизирует его, убирая шум и улучшая качество. Затем изображение выводится на экран или сохраняется на диск.

Примеры выполнения текстового запроса по данному типу алгоритма представлены на рисунках 5 и 6.

Рис. 5. Изображение сгенерированное Stable Diffusion по запросу «The hare is juggling balls», стиль «sai-fantasy art»

Так же следует отметить работу DPM-Solver. Это алгоритм, который позволяет генерировать изображения высокого качества из текстовых описаний. Он основан на использовании нескольких диффузионных моделей и позволяет получать изображения с высоким разрешением и детализацией.

Рис. 6. Изображение созданное Stable Diffusion по запросу «A dog eats a carrot», стиль «cinematic-default»

Код на данном репозитории содержит несколько файлов, которые реализуют различные части алгоритма DPM-Solver, включая класс Sampler, который отвечает за генерацию изображений, и класс DPMSolver, который управляет процессом генерации.

Класс Sampler содержит методы для генерации изображений на основе текстовых описаний, а также для выбора оптимальных параметров для генерации. Класс DPMSolver использует класс Sampler для генерации изображений и управляет процессом обучения модели.

Проанализируем файлы конфигурации. Так, например, файл configs/stable-diffusion/intel/v2-inference-bf16.yaml содержит набор параметров и настроек для модели, включая информацию о количестве слоев, типе активации, функции потерь и других параметрах.

В данном файле описывается конфигурация модели для инференса (процесса генерации изображений) с использованием формата представления чисел BF16 (Brain Floating-Point 16-bit). BF16 — это формат числа с плавающей запятой, который обеспечивает более высокую производительность и эффективность по сравнению с форматом FP32 (стандарт IEEE для чисел с плавающей запятой двойной точности).

Файл содержит настройки для различных этапов обработки изображений, таких как кодирование и декодирование, а также параметры для разных функций активации, таких как ReLU (Rectified Linear Unit), Leaky ReLU и Sigmoid.

Также в файле описываются настройки для регуляризации, которая используется для предотвращения переобучения модели. Регуляризация включает в себя добавление штрафов к функции потерь, чтобы уменьшить сложность модели и предотвратить ее переобучение. Кроме того, в файле описаны параметры для настройки процесса диффузии, которые включают в себя количество шагов диффузии и параметры, связанные с процессом преобразования изображения. Таким образом, этот файл конфигурации описывает все необходимые настройки и параметры для работы модели Stable Diffusion с форматом чисел BF16 при инференсе.

Проанализируем алгоритм диффузии. Диффузия — это процесс распространения чего-либо в среде, например, информации, вещества или тепла. В контексте искусственного интеллекта диффузия может использоваться для создания новых данных или улучшения существующих [4]. Например, в Stable Diffusion используется диффузия для создания изображений из текстовых описаний.

Алгоритм PPLM позволяет управлять процессом диффузии с помощью языковой модели, что позволяет получать более разнообразные и контролируемые результаты.

Файл plms.py содержит классы и методы, которые реализуют алгоритм PPLM, а также дополнительные функции для обработки и анализа данных. В частности, в файле присутствуют следующие классы и методы:

— PPLMController — класс, реализующий алгоритм PPLM.
— PPLMSampler — класс, который реализует процесс диффузии с использованием языковой модели.
— process_batch — функция для обработки данных.
— plot_results — функция для визуализации результатов.
— train_model — функция для обучения модели.

Кроме того, файл содержит переменные и константы, которые определяют параметры алгоритма и другие настройки.

Пример создания изображения на основе текстового запроса:

Сначала, модель получает на вход описание того, что нужно сгенерировать. В нашем случае, это «Собака, которая ест сосиску на пляже в закатное время». Далее код преобразует текст, вычитая лишние символы, с помощью модели CLIP в числовой вектор. Используя набор чисел (шум), модель начинает генерировать изображение. На каждом этапе генерации модель сравнивает полученное изображение с описанием, которое она получила на входе. Если изображение не соответствует описанию, модель вносит изменения в «шум» и генерирует изображение заново. Этот процесс продолжается до тех пор, пока модель не сгенерирует изображение, которое соответствует описанию, таким образом, модель Stable Diffusion генерирует новые изображения на основе заданного описания.

Рис. 7. Изображение созданное Stable Diffusion по запросу «Dog eats sausage on the beach at sunset», стиль «cinematic-default»

На основании проделанной работы можно сделать вывод, что особенностью архитектуры Stable Diffusion является использование механизма диффузии, который позволяет создавать изображения высокого качества и детализации. Также стоит отметить использование автокодирования, которое позволяет сохранять детализацию изображений и улучшать их качество.

В целом, Stable Diffusion представляет собой мощную нейросеть, которая может использоваться для создания разнообразных изображений. Ее архитектура является достаточно гибкой и включает в себя множество слоев и параметров, что позволяет получать высококачественные результаты по заданным запросам.

ЛИТЕРАТУРА

1. Электронный ресурс. Режим доступа: https://4brain.ru/aibasics/deep.php (дата обращения 24.01.2024).
2. Типы нейронных сетей: перцептроны, рекуррентные нейронные сети, сверточные нейронные сети и другие. [Электронный ресурс]. — Режим доступа: https://vc.ru/u/22269-aleksandr-shulepov/675785-tipy-neyronnyh-setey-perceptrony-rekurrentnye-neyronnye-seti-svertochnye-neyronnye-seti-i-drugie. (24.01.2024).
3. Сухань А.А. Генеративно-состязательные нейронные сети в задачах определения трендов-Московский экономический журнал — 2022 — С. 1–8
4. Электронный ресурс. Режим доступа: https://neurohive.io/ru/vidy-nejrosetej/u-net-image-segmentation/ (дата обращения 24.01.2024).
5. Электронный ресурс. Режим доступа: https://ya.ru/images/search?from=tabbar&img_url=https%3A%2F%2Fb2633864.smushcdn.com%2F2633864%2Fwp-content%2Fuploads%2F2021%2F11%2Fu-net_training_image_segmentation_models_in_pytorch_header.png%3Flossy%3D2%26strip%3D1%26webp%3D1&lr=213&pos=1&rpt=simage&text=U-Net (дата обращения 24.01.2024).
6. Электронный ресурс. Режим доступа: https://neiroseti.tech/neyroseti/resnet-obzor/ (дата обращения 24.01.2024).
7. Электронный ресурс. Режим доступа: https://ya.ru/images/search?from=tabbar&img_url=https%3A%2F%2Fwww.mdpi.com%2Fremotesensing%2Fremotesensing-11-01117%2Farticle_deploy%2Fhtml%2Fimages%2Fremotesensing-11-01117-g006.png&lr=213&pos=14&rpt=simage&text=ResNet (дата обращения 24.01.2024).
8. Электронный ресурс. Режим доступа: https://habr.com/ru/articles/539312/ (дата обращения 24.01.2024).
9. Электронный ресурс. Режим доступа: https://www.evogeek.ru/articles/66374/ (дата обращения 24.01.2024).

Журнал «Современная наука: актуальные проблемы теории и практики»