УДК 004

Создание реалистичных моделей лица с помощью Refined Stable Diffusion Model

Беликов Андрей Геннадьевич
Приамурский государственный университет имени Шолом-Алейхема
Студент

Аннотация
В данной статье исследуется методология, лежащая в основе создания модели лица.
Ключевые слова: Stable Diffusion, ИИ, лицо

Creating realistic facial models using Refined Stable Diffusion Model

Belikov Andrey Gennadievich
Sholom-Aleichem Priamursky State University
Student

Abstract
This article examines the methodology underlying the creation of a face model.
Keywords: Stable Diffusion, AI, face

В данной статье показан процесс создания реалистичных моделей лица с помощью Refined Stable Diffusion Model. В результате работы была создана реалистичная модель лица нейросетью Stable Diffusion.

Цель данной статьи создание реалистичной модели лица с помощью Refined Stable Diffusion Model.

Для создания проекта была рассмотрена статья А. М. Мартыненко и С. В. Васильев, в которой они анализируют нейронную сеть «stable diffusion» для генерации фотографий [1]. В статье А. П. Лосева, Д. А. Поленовой, Е. И. Тумановой, описывается возможность оценки качества компрессии изображений при помощи модели нейронной сети stable diffusion [2]. Была рассмотрена статья Г. А. Урванцева, К. Т. Шариповой, А. П. Маринской в которой был произведен анализ перспектив применения мультимодальных нейросетевых технологий в современном медиамаркетинге на примере stable diffusion [3].

Stable Diffusion - это искусственный интеллект, с преобразованием текста в изображение, выпущенная в 2022 году. В основном он используется для создания подробных изображений, основанных на текстовых описаниях, его также можно применять для других задач, таких как ввод, вывод изображения и генерация переводов между изображениями, управляемых текстовой подсказкой.

Первым делом необходимо показать сам процесс создания модели (Рисунок 1).

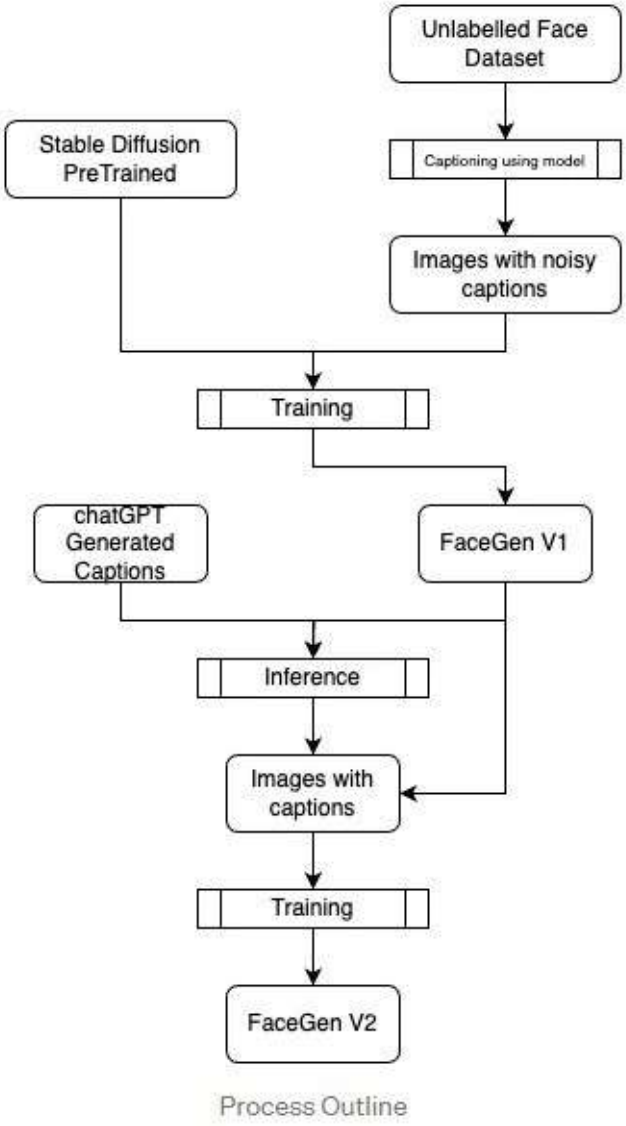

Рисунок 1. Процесс создания модели

Данная модель генерации была обучена на более чем 52 000 наборах данных о лицах, состоящих из разнообразного набора изображений лиц, охватывающего широкий спектр черт, выражений и демографических данных. Модель V1 была улучшена с помощью дополнительного ввода данных, например, об цвете волос, чтобы убрать неточности при генерации нового изображения, например, такие как 6 пальцев на руках.

Ниже представлены несколько поколений модели (версия 2). Как видно, разнообразие и реалистичность, которые может создать модель отличаются друг от друга. Некоторые настройки можно дополнительно прописать, используя соответствующие подсказки, такие как “фиолетовые волосы”, “короткая стрижка” (Рисунок 2).

Рисунок 2. Примеры генерации лиц

Используемые скрипты были взяты из Kohya Stable Diffusion Trainer.

Компоненты Stable Diffusion, которые используются для обработки запроса — это система, состоящая из множества компонентов, а именно она содержит компонент понимания текста, преобразующий текстовую информацию в цифровой вид, который передаёт заложенный в текст смысл (Рисунок 3).

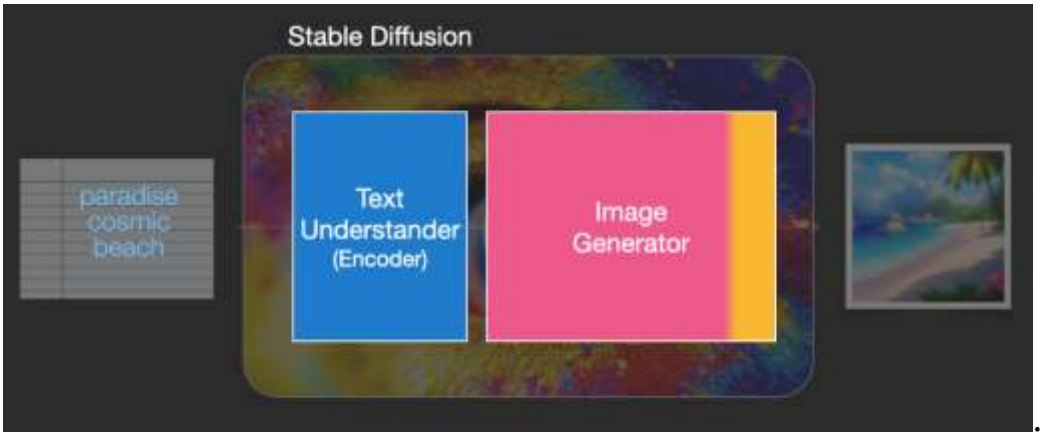

Рисунок 3. Работа нейросети

Одним из элементов, которые находятся в модели – это кодировщик текста. Он является специальной языковой моделью Transformer (технически её можно описать как текстовый кодировщик модели CLIP). Она получает на входе текст и выдаёт на выходе список чисел (вектор), описывающий каждое

слово/токен в тексте. Далее эта информация передаётся генератору изображений, который состоит из двух компонентов (Рисунок 4).

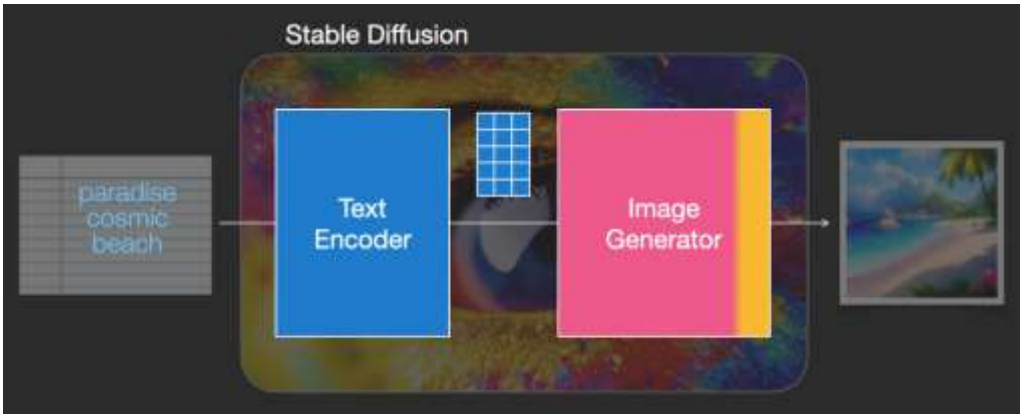

Рисунок 4. Работа нейросети

Генератор изображений выполняет два этапа:

1- Создание информации изображения.

Этот компонент выполняется в несколько шагов (step), генерируя информацию изображения. Это параметр steps в интерфейсах и библиотеках Stable Diffusion, который часто по умолчанию имеет значение 50 или 100.

Этап создания информации изображения действует полностью в пространстве информации изображения (или в скрытом пространстве). Подробнее о том, что это значит, мы расскажем ниже. Это свойство ускоряет работу по сравнению с предыдущими моделями диффузии, работавшими в пространстве пикселей. Этот компонент состоит из нейросети UNet и алгоритма планирования.

Слово «диффузия» (diffusion) описывает происходящее в этом компоненте. Это пошаговая обработка информации, приводящая в конечном итоге к генерации высококачественного изображения (при помощи следующего компонента — декодера изображений).

2- Декодер изображений.

Декодер изображений рисует картину на основе информации, которую он получил на этапе создания информации. Он выполняется только один раз в конце процесса и создаёт готовое пиксельное изображение (Рисунок 5-6).

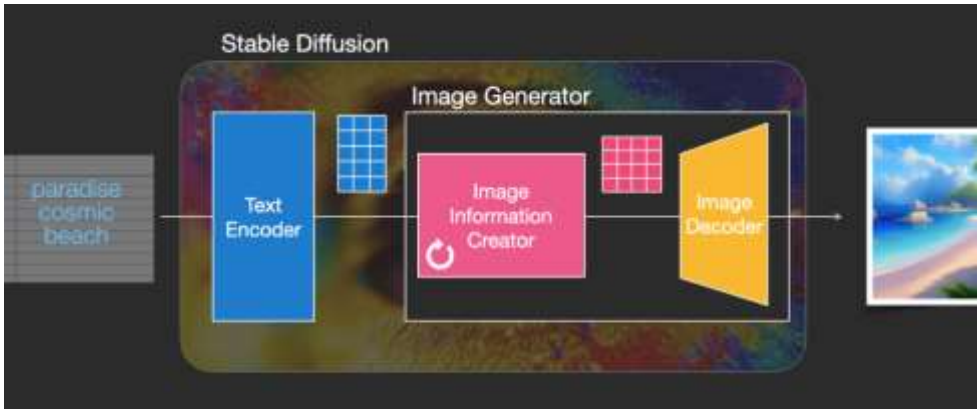

Рисунок 5. Работа нейросети

Рисунок 6. Работа нейросети

ClipText элемент для кодирования текста.

Выходные данные: 77 векторов эмбеддингов токенов, каждый в 768 измерениях.

UNet + Scheduler для постепенной обработки/диффузии информации в пространстве информации (скрытом пространстве).

Входные данные: эмбеддинги текста и исходный многомерный массив (структурированные списки чисел, также называемые тензором), состоящий из шума.

Выходные данные: массив обработанной информации

Декодер автокодировщика, рисующий готовое изображение при помощи массива обработанной информации.

Входные данные: массив обработанной информации (размеры: (4,64,64))

Выходные данные: готовое изображение (размеры: (3, 512, 512) — (красный/зелёный/синий, ширина, высота)) (рисунок 7).

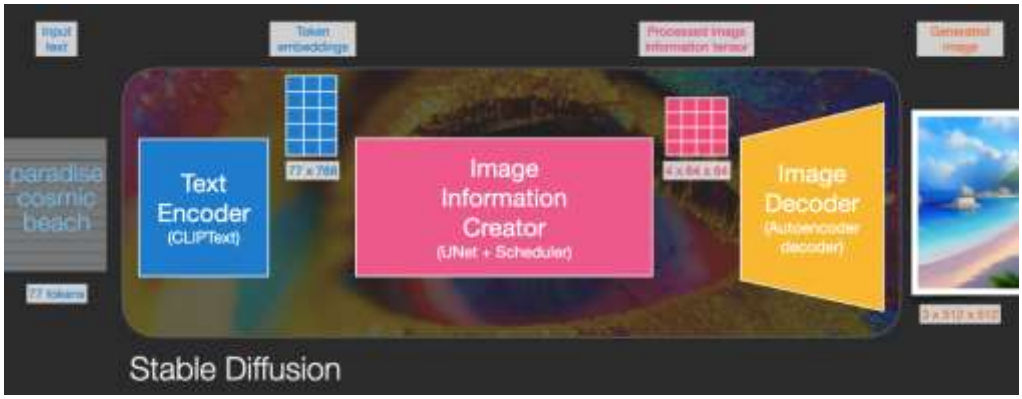

Рисунок 7. Работа нейросети

Процесс работы диффузии и основная идея генерации изображений при помощи диффузионной модели использует тот факт, что имеются мощные модели компьютерного зрения. Если им передать достаточно большой массив данных, эти модели могут обучаться сложным операциям. Диффузионные модели подходят к задаче генерации изображений, формулируя задачу следующим образом:

Допустим, у нас есть изображение, сделаем первый шаг, добавив в него немного шума (Рисунок 8).

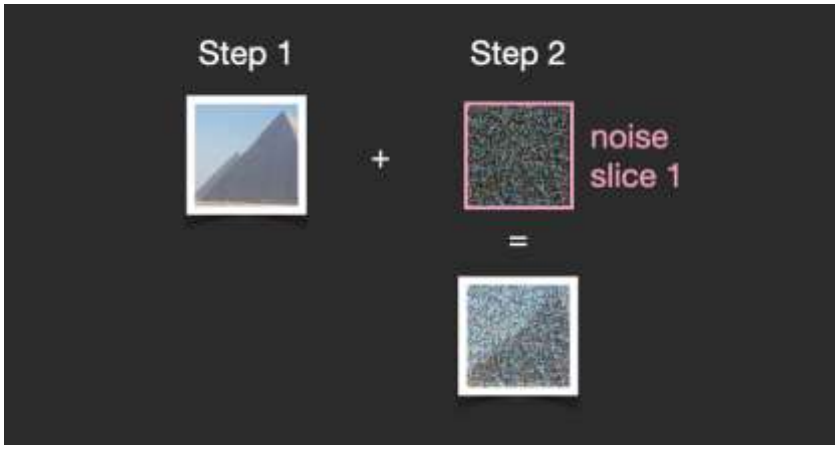

Рисунок 8. Добавление шума

Назовём «срез» (slice) добавленного нами шума «noise slice 1». Сделаем ещё один шаг, добавив к шумному изображению ещё шума («noise slice 2») (Рисунок 9).

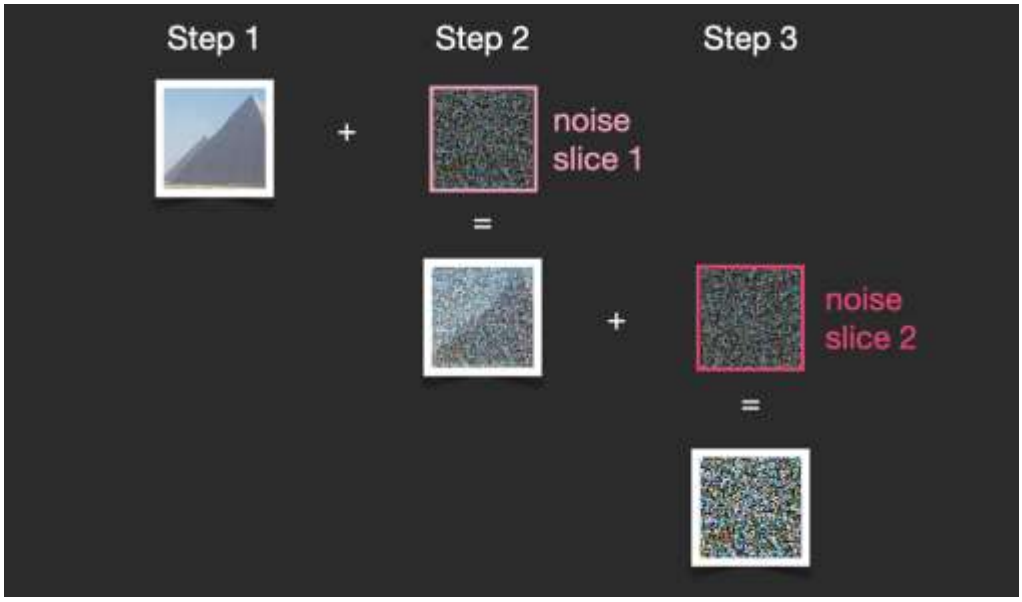

Рисунок 9. Добавление шума

На этом этапе изображение полностью состоит из шума. Теперь давайте возьмём их в качестве примеров для обучения нейронной сети компьютерного зрения. Имея номер шага и изображение, хотим, чтобы она спрогнозировала, сколько шума было добавлено на предыдущем шаге (Рисунок 10).

Рисунок 10. Добавление шума

На данном этапе есть возможность управлять тем, сколько шума добавляется к изображению, поэтому можно распределить его на десятки шагов, создав десятки примеров для обучения на каждое изображение для всех изображений в обучающем массиве данных (Рисунок 11).

Рисунок 11. Обучение нейросети

После того, как эта сеть прогнозирования шума начнёт работать правильно, она, по сути, сможет рисовать картины, удаляя шум на протяжении множества шагов.

Примечание: это небольшое упрощение алгоритма диффузии. На ресурсах по ссылкам в конце статьи представлено более подробное математическое описание.

В данной статье показан процесс создания реалистичных моделей лица с помощью Refined Stable Diffusion Model. В результате работы была создана реалистичная модель лица нейросетью Stable Diffusion.

Библиографический список

1. Мартыненко А. М., Васильев С. В. Анализ нейронных сетей «stable diffusion» для генерации фотографий, по преобразованию текста в изображение / В сборнике: Донецкие чтения 2022: образование, наука, инновации, культура и вызовы современности. Материалы VII Международной научной конференции, посвящённой 85-летию Донецкого национального университета. Под общей редакцией С.В. Беспаловой. Донецк, 2022. С. 265-267.
2. Лосев А. П., Поленова Д. А., Туманова Е. И. Оценка качества компрессии изображений при помощи модели нейронной сети stable diffusion / В сборнике: Подготовка профессиональных кадров в магистратуре для цифровой экономики (ПКМ-2022). Сборник лучших докладов Всероссийской научно-технической и научно-методической конференции магистрантов и их руководителей. Сост. Н.Н. Иванов. Санкт-Петербург, 2023. С. 92-96.
3. Урванцев Г. А., Шарипова К. Т., Маринская А. П. Анализ перспектив применения мультимодальных нейросетевых технологий в современном медиамаркетинге на примере stable diffusion / В сборнике: Вестник факультета социальных цифровых технологий санкт-петербургского государственного университета телекоммуникаций им. проф. м. а. Бонч-бруевича. сборник научно-теоретических статей. Санкт-Петербург, 2022. С. 134-139.