УДК 004.93 (687.01)

ГЕНЕРАЦИЯ ИЗОБРАЖЕНИЙ ОДЕЖДЫ ПРИ ПОМОЩИ МОДЕЛИ KANDINSKY

А. О. Яблонская[1)], А. Э. Малевич[2)]

[1)]Белорусский государственный университет, Беларусь, Минск, anna.yablonskaya2002@gmail.com
[2)]Белорусский государственный университет, Беларусь, Минск, malevich@bsu.by

В данной статье рассматривается разработка генеративной модели для создания изображений одежды. В качестве модели выбрана предобученная модель Kandinsky. Представлена общая информация об используемой модели, а также результаты её обучения для поставленной задачи.

Ключевые слова: генерация изображений; генеративные модели; модель Kandinsky; архитектура модели; обучение модели; гиперпараметры.

GENERATION OF CLOTHING IMAGES USING THE KANDINSKY MODEL

A. O. Yablonskaya[1)], A. E. Malevich[2)]

[1)]Belarussian state university, Belarus, Minsk, anna.yablonskaya2002@gmail.com
[2)]Belarussian state university, Belarus, Minsk, malevich@bsu.by

This article discusses the development of a generative model for creating clothing images. Kandinsky model is chosen as a pretrained model. The article contains general information about the model and the result of its training for the task.

Keywords: image generation; generative models; Kandinsky model; model architecture; model training; hyperparameters.

Введение

В современном мире мода и стиль играют важную роль в жизни людей, определяя не только внешний вид, но и самоидентификацию. Системы онлайн-стилистов, использующие методы машинного обучения, предлагают инновационные решения для персонализированного подхода в выборе одежды и аксессуаров. Эти технологии способны анализировать

предпочтения пользователей, комбинацию цветов и форм, а также учитывать актуальные тренды, что значительно упрощает процесс поиска подходящих предметов одежды и создание стильных образов.

Одной из подзадач в создании онлайн-стилиста является генерация новых изображений на основе описания предмета и исходной картинки. Генерация новых изображений даёт возможность клиентам видеть их идеи и пожелания в визуальной форме, создавая уникальные предложения для каждого пользователя.

В данной статье будет описан процесс построения генеративной модели для создания изображений одежды.

Генеративные модели

Генеративные модели — класс статистических алгоритмов машинного обучения. Эти модели способны генерировать новые экземпляры данных, которые напоминают обучающие данные. Генеративные модели изучают распределение данных входного обучающего набора с целью генерировать новые точки данных, которые напоминают исходный обучающий набор. Это означает, что такие модели способны понимать и воспроизводить нюансы данных. Генеративные модели нашли широкое применение, начиная от генерации изображений и заканчивая пониманием и синтезом естественного языка.

Генеративные модели работают, фиксируя совместное распределение вероятностей между наблюдаемыми переменными (фактической точкой данных) и скрытыми переменными (возможными или вероятными точками данных). Это совместное распределение затем используется для генерации новых данных. Основной фокус этих моделей — понять базовую структуру и распределение данных. [1]

Модель Kandinsky

Модель Кандинский (или Kandinsky) является примером генеративной модели, разработанной для создания изображений на основе текстовых описаний. Она основана на архитектуре трансформеров и предназначена для генерации высококачественного визуального контента. Существует различные версии данной модели. Для решения задачи применялась версия Kandinsky 2.1 (рис. 1).

Кандинский 2.1 унаследовал лучшие практики от Dall-E 2 и скрытой диффузии, но привнёс некоторые новые идеи [2].

Архитектура модели:

• Text encoder: XLM-Roberta-Large-Vit-L-14 — это модель, объединяющая в себе технологии обработки естественного языкаи компьютерного

зрения. Она основана на архитектуре XLM-RoBERTa, представляющей собой многоязычную предобученную модель трансформеров для понимания текста, и ViT (Vision Transformer) для обработки изображений.

• Diffusion Image Prior — это метод, использующий диффузионные модели для генерации и восстановления изображений, основанный на принципах диффузии и статистического обучения.

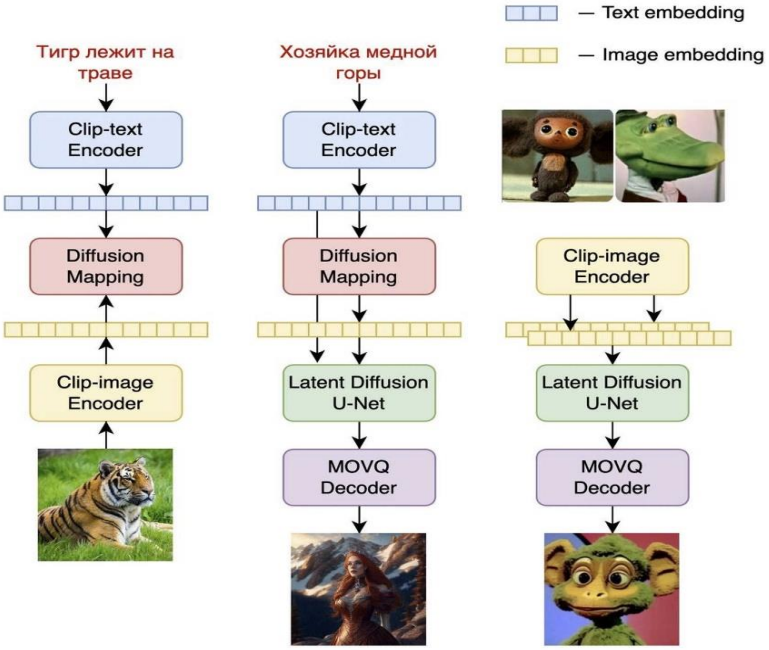

Рис.1. Архитектура модели Kandinsky

• CLIP image encoder (ViT-L/14 — это архитектура для обработки изображений, основанная на трансформерах) Данная модель представляет собой расширенный вариант архитектуры ViT. Энкодер возвращает вектор, соответствующий изображению.

• Latent Diffusion U-Net — архитектура, используемая в контексте диффузионных моделей, которая сочетает в себе преимущества U-Net и латентного представления данных. Это позволяет более эффективно обрабатывать изображения, снижая вычислительную нагрузку и увеличивая качество генерации

• MoVQ encoder/decoder — подход в области кодирования и декодирования, используемый для эффективной обработки многоязычных и многовидовых данных. В общем случае, MoVQ включает в себя encoder и decoderб которые работают совместно для достижения эффективной передачи и восстановления информации.

Кандинский 2.1 был обучен на обширном наборе данных изображений и текста LAION HighRes и доработан на внутренних данных Сбербанка. Модель image prior обучалась на датасете LAION Improved Aesthetics, после чего прошла файнтюнинг на LAION HighRes. Основная Text2Image диффузионная модель была обучена на 170 миллионах пар

«текст-изображение» из датасета LAION HighRes, при этом изображения имели разрешение не менее 768x768 пикселей. На этапе файнтюнинга был использован отдельно собранный датасет из 2 миллионов качественных изображений в высоком разрешении с описаниями, включая COYO, anime, landmarks_russia и другие источники.

Обучение модели

Данные

Для выполнения поставленной задачи генерации изображений одежды в первую очередь необходимо подготовить датасет, содержащий изображения и описание данных изображений.

В качестве датасета был выбран FashionGen датасет, содержащий 32 528 изображений предметов одежды и их описания.

Обучение

Далее для применения модели к поставленной задачи, её необходимо было дообучить на новых примерах. При обучении были применены следующие параметры:

• num_epochs (количество эпох) = 200;

• batch size (колчество экземпляров в одном баче) = 32;

• clip_image_size (размер входных изображений, который модель принимает для обработки, используется на уровне) = 224;

• diffusion steps (количество интераций, через который проходит процесс генерации) = 500;

• learning rate (гиперпараметр оптимизации, контролирующий скорость обновления весов модели) = 0.0005.

После обучения, модель можно применять, как для генерации изображений по текстовому запросу, так и на основе каких-либо изображений.

Пример. Из исходного изображения получить изображение по запросу: черная кожаная мини-юбка (см. рис. 2).

1 *2*

Рис. 2. Генерация изображений: 1 – исходное, 2 – сгенерированное

Заключение

В заключение, использование технологий машинного обучения для создания онлайн-стилистов представляет собой революционный шаг в сфере моды и стиля. Генерация изображений на основе описания и исходных картинок не только облегчает пользователям процесс подбора одежды, но и предоставляет уникальные возможности для самовыражения. В статье была представлена архитектура генеративной модели Kandinsky, на основе которой можно построить модель для создания новых изображений по запросам пользователя.

Библиографические ссылки

1. *Ivan Belcic.* What is a generative model?, 2024. [Online]. URL: https://www.ibm.com/think/topics/generative-model (date of access: 10.04.2025).

2. *Кузнецов А.* Kandinsky 2.1 или Когда +0.1 значит очень много, 2023. [Online]. URL: https://habr.com/ru/companies/sberbank/articles/725282/ (date of access: 10.04.2025).