Научная статья
УДК 004.932.72'1
ББК 32.818.1
И 88
DOI: 10.53598/2410-3225-2022-4-311-66-70

Использование диффузионных моделей для разработки приложений, генерирующих изображения на основе текстовых запросов

(*Рецензирована*)

Альфира Менлигуловна Кумратова[1], Мариям Адильовна Борлакова[2], Виктор Евгеньевич Сайкинов[3], Ирина Евгеньевна Когай[4]

[1, 3, 4] *Кубанский государственный аграрный университет*, *Краснодар*, *Россия*
[2] *Институт информационных технологий Северо-Кавказской государственной академии*, *Черкесск*, *Россия*, *borlakova_mar@mail.ru*
[1] *kumratova.a@edu.kubsau.ru*
[3] *sajkinov2014@yandex.ru*
[4] *ikogai.me@gmail.com*

Аннотация. *Целью данной статьи является рассмотрение диффузионных моделей, а также исследование возможностей применения данных нейронных сетей для разработки приложений, генерирующих изображения на основе текстовых запросов, сравнение диффузионных моделей, существующих на сегодняшний день. Диффузионные модели – подкатегория глубоких генеративных моделей, которые состоят из этапов прямой и обратной диффузии, генерируют данные, аналогичные тем, на которых они обучаются. Разные архитектуры диффузионных нейронных систем могут быть использованы как для генерации изображения на основании текстовых запросов, так и для преобразования существующих изображений. В перспективе развития диффузионных нейронных сетей можно предположить, что их использование существенно облегчит работу дизайнеров на производстве.*

Ключевые слова: *нейронные сети, диффузионные модели, генерация изображений, текстовый запрос, диффузия, архитектура диффузионных нейронных систем*

Original Research Paper

Using diffusion models to develop applications that generate images based on text queries

Alfira M. Kumratova[1], Mariyam A. Borlakova[2], Viktor E. Saykinov[3], Irina E. Kogay[4]

[1, 3, 4] *Kuban State Agrarian University*, *Krasnodar*, *Russia*
[2] *Institute of Information Technologies of the North-Caucasus State Academy*, *Cherkessk*, *Russia*, *borlakova_mar@mail.ru*
[1] *kumratova.a@edu.kubsau.ru*
[3] *sajkinov2014@yandex.ru*
[4] *ikogai.me@gmail.com*

Abstract. *The purpose of this article is to consider diffusion models, as well as to explore the possibilities of using neural network data to develop applications that generate images based on text queries, and to compare the diffusion models that exist today. Diffusion models are a subcategory of deep generative models, which consist of stages of forward and reverse diffusion, generate data similar to those they use to train. Different architectures of diffusion neural systems can be used both to generate images based on text queries and to transform existing images. In the perspective of the de-*

velopment of diffusion neural networks, we assume that their use will significantly facilitate the work of designers in production.

Keywords: *neural networks*, *diffusion models*, *image generation*, *text query*, *diffusion*, *architecture of diffusion neural systems*

Диффузионные модели – подкатегория глубоких генеративных моделей, которые состоят из этапов прямой и обратной диффузии, генерируют данные, аналогичные тем, на которых они обучаются.

Вопросам распознавания образов и изображений на основе нейронных сетей посвящены труды многих авторов [1–5].

Диффузионные модели обусловлены неравновесной термодинамикой. Они определяют марковскую цепь шагов диффузии, чтобы медленно добавлять случайный шум к данным, а затем учатся обращать процесс диффузии вспять для создания желаемых выборок данных из шума [6]. В отличие от вариационных автоэнкодеров (VAE) или моделей потока, модели диффузии обучены с помощью фиксированной процедуры, а скрытая переменная имеет высокую размерность (такую же, как исходные данные) [7]. Диффузионные модели состоят из следующих этапов:

1.Прямая диффузия (Forward Diffusion) – искажение обучающих данных путем поэтапного добавления гауссовского шума и стирания деталей, пока данные не будут преобразованы в чистый шум [8];

2.Параметризованный реверс – обучение нейронной сети для того, чтобы обратить процесс искажения вспять, то есть восстановить исходное изображение, синтезируя чистый шум путем постепенного снижения шума до тех пор, пока не будет получен чистый образец.

Вышеописанные процессы занимают много времени и являются достаточно медленными. Это связано с тем, что этапы требуют последовательного повторения тысячи шагов добавления и устранения шума.

На сегодняшний день существует несколько видов архитектур диффузионных нейронных сетей, среди них:

1. Stable Diffusion – состоит из автоэнкодера, блока U-Net и декодера. Рассмотрим указанные части Stable Diffusion подробнее:

– Автоэнкодер редуцирует данные в латентное пространство до меньшего размера;

– Блок U-Net получает зашумленный образец в латентном пространстве, сжимает его и декодирует обратно с меньшим уровнем шума. Для построения ожидаемого представления зашумленного образца U-Net использует вычисленный на выходе остаток шума;

– Кодировщик текста обрабатывает его, преобразуя текстовый запрос в место для встраивания.

Стоит отметить, что архитектура Stable Diffusion практически не имеет ограничений на контент, который она способная генерировать, в связи с чем она может генерировать фотографии людей или изображения, похожие на работы художников, которые не давали согласия на использование своих материалов [9, 10] (в соответствии с рисунком 1).

2. Архитектура Kandinsky 2.0 – первая российская мультиязычная диффузионная модель, в основе которой лежит улучшенный подход Latent Diffusion. Модель понимает такие языки, как: английский, русский, монгольский, фарси, немецкий, французский и другие. Kandinsky 2.0 содержит в себе два мультилингвальных текстовых энкодера (mCLIP-XLMR – 560 миллионов параметров и mT5-encoder-small – 146 миллионов параметров) и U-Net Блок и декодер.

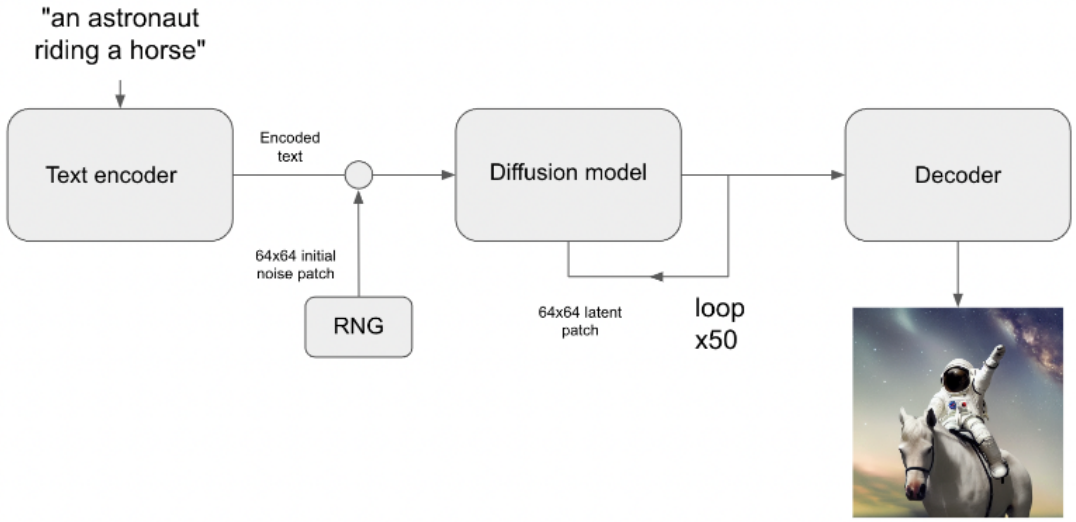

Рис. 1. Схема для генерации изображения на основе диффузионных моделей

Fig. 1. Diagram for generating an image based on diffusion models

Kandinsky 2.0 имеет несколько способов применения:

1) Text to image – преобразование текстового запроса (prompt) в изображение;

2) Inpainting – процесс замены специально поврежденных или закрашенных фрагментов изображений новыми сгенерированными частями;

3) Image to image – процесс создания нового изображения на базе введенных изображений на основании нахождения соответствия между входным и выходным изображениями (в соответствии с рисунком 2).

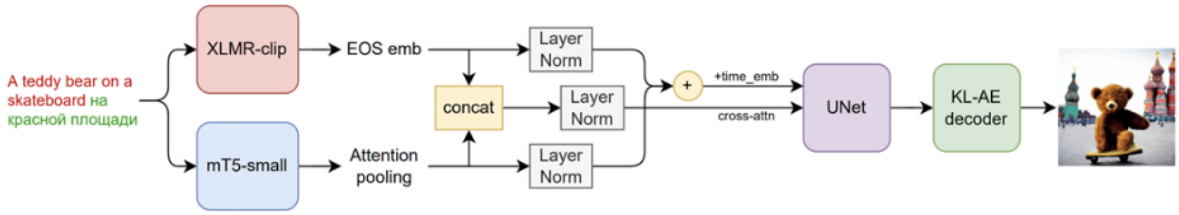

Рис. 2. Схема создания изображения на основе текстовых запросов

Fig. 2. A scheme to create an image from a text query

Используя подход *text to image*, можно создать изображение на основе текстового запроса:

```
pip install "git+https://github.com/ai-forever/Kandinsky-2.0.git"

from kandinsky2 import get_kandinsky2
model = get_kandinsky2('cuda', task_type='text2img')

images = model.generate_text2img('космосвнутричайнойчашки', batch_size=4, h=512, w=512, num_steps=75, denoised_type='dynamic_threshold', dynamic_threshold_v=99.5, sampler='ddim_sampler', ddim_eta=0.01, guidance_scale=10)

images [3]
```

На рисунке 3 представлен результат работы по созданию изображения на основании текстового запроса «космос внутри чайной чашки»:

Рис. 3. Изображение на основе текстового запроса «космос внутри чайной чашки»

Fig. 3. Image based on text query “space inside tea cup”

3. Disco Diffusion – диффузионная модель, управляемая Clip, генерирующая изображения на основе текстового запроса. Эта модель появилась из Google Colab (Катрин Кроусон) и ее отлаженной диффузионной модели, которую затем стали развивать и поддерживать многие другие разработчики из разных сообществ.

Отметим, что многие коммерческие проекты используют различные архитектуры для создания приложений с целью извлечения прибыли. Среди самых известных: DALLE 2, Midjourney, Lensa.

Данные приложения генерируют изображения на основе текстовых запросов, а также создают новые изображения на базе загруженных ранее.

На основании тезисов, перечисленных в данной статье, можно сделать вывод, что разные архитектуры диффузионных нейронных систем могут быть использованы как для генерации изображении на основании текстовых запросов, так и для преобразования существующих изображений. В перспективе развития диффузионных нейронных сетей можно предположить, что их использование существенно облегчит работу дизайнеров на производстве.

Примечания

1. Бережнов Н.И., Сирота А.А. Универсальный алгоритм улучшения изображений с использованием глубоких нейронных сетей // Вестник Воронежского государственного университета. Сер.: Системный анализ и информационные технологии. 2022. № 2. С. 81–92.

2. Войнов Д.М., Ковалев В.А. Устойчивость нейронных сетей к состязательным атакам при распознавании биомедицинских изображений // Журнал Белорусского государственного университета. Математика. Информатика. 2020. № 3. С. 60–72.

3. Генерация текстур с использованием сверточных нейронных сетей / О.М. Бакунова, А.М. Бакунов, И.Л. Калитеня [и др.] // Web of Scholar. 2018. Т. 1, № 4 (22). С. 16–18.

4. Малыгина А.Д., Соков Б.Б. Сверточные нейронные сети в задаче генерации изображений // Аллея науки. 2019. Т. 4, № 1 (28). С. 542–547.

5. Кадомский А.А., Сабинин О.Ю. Исследование возможности стилизации материалов

трехмерных моделей, основанной на аппарате искусственных нейронных сетей // Theoretical & Applied Science. 2021. № 2 (94). С. 165–176.

6. Siddiqui J. Rafid. Diffusion Models Made Easy. PhD, 2022. URL: https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da (дата обращения: 23.11.2022).

7. Worksolutions. Нейродайджест: главное из области машинного обучения за июль 2021. 2 августа 2021. URL: https://habr.com/ru/post/570912/ (дата обращения: 23.11.2022).

8. Lilian Weng What are Diffusion Models? 11 июля 2021. URL: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/ (дата обращения: 22.11.2022).

9. Ryan O'connor Introduction to Diffusion Models for Machine Learning. 12 мая 2022. URL: https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/ (дата обращения: 21.11.2022).

10. Rombach R., Blattman A., Lorenz D. High-Resolution Image Synthesis with Latent Diffusion Models. URL: https://arxiv.org/pdf/2112.10752.pdf (дата обращения: 11.11.2022).

References

1. Berezhnov N.I., Sirota A.A. Universal image enhancement algorithm using deep neural networks // Bulletin of the Voronezh State University. Ser.: System Analysis and Information Technologies. 2022. No. 2. P. 81–92.

2. Voynov D.M., Kovalev V.A. Resistance of neural networks to competitive attacks when recognizing biomedical images // Journal of the Belarusian State University. Mathematics. Computer Science. 2020. No. 3. P. 60–72.

3. Texture generation using precise neural networks / O.M. Bakunova, A.M. Bakunov, I.L. Kalitenya [et al.] // Web-scientist. 2018. Vol. 1, No. 4 (22). P. 16–18.

4. Malygina A.D., Sokov B.B. Convolutional neural networks-ti in the problem of image generation // Alley of Science. 2019. Vol. 4, No. 1 (28). P. 542–547.

5. Kadomsky A.A., Sabinin O.Yu. Investigation of the possibility of stylization of materials of three-dimensional models based on the apparatus of artificial neural networks // Theoretical & Applied Sciences. 2021. No. 2 (94). P. 165–176.

6. Siddiqui J. Rafid. Diffusion Models Made Easy. PhD, 2022. URL: https://towardsdatascience.com/diffusion-models-made-easy-8414298ce4da (access date: 23.11.2022).

7. Worksolutions. Neurodigest: the main thing from the field of machine learning for July 2021. August 2, 2021. URL: https://habr.com/ru/post/570912/ (access date: 23.11.2022).

8. Lilian Weng What are Diffusion Models? July 11, 2021. URL: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/ (access date: 22.11.2022).

9. Ryan O'connor Introduction to Diffusion Models for Machine Learning. May 12, 2022. URL: https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/ (access date: 21.11.2022).

10. Rombach R., Blattman A., Lorenz D. High-Resolution Image Synthesis with Latent Diffusion Models. URL: https://arxiv.org/pdf/2112.10752.pdf (access date: 11.11.2022).

Авторы заявляют об отсутствии конфликта интересов.

The authors declare no conflicts of interests.

Статья поступила в редакцию 25.11.2022; *одобрена после рецензирования 15.12.2022*; *принята к публикации 16.12.2022*.

The article was submitted 25.11.2022; *approved after reviewing 15.12.2022*; *accepted for publication 16.12.2022*.