УДК 004.932 **Технические науки**

Левин Артем Олегович, *студент-магистр, Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»*

Белов Юрий Сергеевич, *к.ф. -м.н., доцент, Калужский филиал ФГБОУ ВО «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)»*

ПРИМЕНЕНИЕ МОДЕЛИ НИЗКОРАНГОВОЙ АДАПТАЦИИ ДЛЯ ГЕНЕРАЦИИ ИЗОБРАЖЕНИЙ ПО ТЕКСТОВОМУ ОПИСАНИЮ СОВМЕСТНО С ДИФФУЗИОННЫМИ МОДЕЛЯМИ

Аннотация: В данной статье рассматривается использование метода LoRA (LowRankAdaptation) в сочетании с диффузионными моделями для генерации изображений на основе текстовых описаний. Данный метод позволяет сократить вычислительную сложность и ускорить процесс обучения путем применения низкоранговых слоев, полученных с использованием SVD (Singular Value Decomposition) или слоев сжатия входных данных (Input Compression Layers). Эти слои заменяют полносвязные слои в архитектуре диффузионной модели, что позволяет уменьшить количество параметров модели, сохраняя ее обучаемость и улучшая качество генерируемых изображений. Также возможно реализовать модель LoRA, для совместной работы с главной диффузионной моделью.

Ключевые слова: Генерация изображений, диффузионные модели, низкоранговая адаптация, LoRA, T2I.

Abstract: This article describes the usage of LoRA (LowRankAdaptation) method in combination with diffusion models to generate images based on text descriptions. This method allows to reduce computational complexity and speed up

the learning process by using low-rank layers obtained using SVD (Singular Value Decomposition) or Input Compression Layers. These layers replace dense layers in the architecture of the diffusion model, which makes it possible to reduce the number of model parameters, while maintaining its trainability and improving the quality of generated images. It is also possible to implement a LoRA model to work in conjunction with the main diffusion model.

Keywords: Image synthesis, diffusion models, low-rank adaptation, LoRA, T2I.

Введение. Современные системы генерации изображений на основе текстовых описаний являются важным направлением в области искусственного интеллекта и компьютерного зрения. Одной из наиболее перспективных и быстро развивающихся технологий в этой области являются диффузионные модели. Они позволяют генерировать качественные изображения на основе текстовых описаний, используя нейронные сети и алгоритмы диффузии. Однако, при работе с большими объемами данных, применение диффузионных моделей сталкивается с проблемой вычислительной сложности. Одним из возможных решений этой проблемы может быть применение низкоранговой адаптации LowRankAdaptation (LoRA), позволяющая уменьшить количество параметров и сократить время обучения моделей.

Общая информация. LoRA (LowRankAdaptation) - это метод снижения размерности, который позволяет уменьшить количество параметров в нейронной сети путем замены полносвязных слоев на низкоранговые слои. Полносвязные слои имеют большое количество параметров, что может приводить к переобучению модели и высокой вычислительной сложности. Низкоранговые слои, в свою очередь, имеют значительно меньшее количество параметров, что позволяет снизить риск переобучения и ускорить процесс непосредственного обучения [1].

Процесс обучения нейронной сети с применением LoRA состоит из нескольких шагов. Сначала проводится обычное обучение нейронной сети с

использованием полносвязных слоев. Затем, на основе полученных параметров, производится аппроксимация полносвязных слоев низкоранговыми слоями. В качестве низкоранговых слоев можно использовать, например, слои сингулярного разложения (SVD) или слои сжатия входа (Input Compression Layers). После этого производится дообучение модели с использованием полученных низкоранговых слоев [2].

Преимущества применения LoRA заключаются в сокращении количества параметров и ускорении процесса обучения, что особенно актуально для диффузионных моделей.

Диффузионные модели являются вероятностными моделями, которые используются для изучения распределения p(x) заданного набора данных. Целью этих моделей является удаление шума из нормально распределенной переменной. Этот процесс соответствует обратному процессу обучения фиксированной цепи Маркова длины T. В контексте синтеза изображений эти модели используют завышенную вариационную нижнюю границу p(x) для отражения результатов шумоподавления. Для этого используется последовательность шумоподавляющих автокодировщиков $\epsilon_\theta(x_t, t); t = 1 \dots T$, каждый из которых имеет одинаковый вес [3]. Каждый автокодировщик обучен предсказывать исходную версию входных данных, обозначаемую как x_0, с учетом зашумленной версии входных данных, обозначаемой как x_t, на каждом шаге t. Таким образом, можно упростить соответствующую задачу до

$$L_{DM} = E_{x,\epsilon\sim N(0,1),t}[||\epsilon - \epsilon_\theta(x_t, t)||_2^2],$$

где t равномерно выбирается из {1, . . . , T}.

Как и другие генеративные модели, диффузионные модели, позволяют моделировать условные распределения вида p(z|y). Чтобы добиться этого, можно использовать автокодировщик с условным шумоподавлением θ(zt, t, y), который контролирует процесс синтеза с учетом входных данных Y. Эти данные могут представлять собой текст, семантические карты или другие типы данных, которые могут быть преобразованы изображениями или текстом в

изображения [4].

Обобщая, диффузионные модели позволяют сгенерировать изображение по текстовому описанию путем последовательного изменения шума в пространстве пикселей. Шум постепенно "растекается" по пространству, образуя изображение с заданными характеристиками. Однако, при работе с большими объемами данных, обучение диффузионных моделей может занимать много времени и требовать больших вычислительных ресурсов [5].

Применение LoRA совместно с диффузионными моделями позволяет существенно снизить вычислительную сложность и ускорить процесс обучения. Низкоранговые слои, полученные с помощью LoRA, могут быть использованы в качестве замены полносвязным слоям в архитектуре диффузионной модели [6]. Это позволяет снизить количество параметров и ускорить процесс обучения, что особенно актуально при работе с большими объемами данных (Рис. 1).

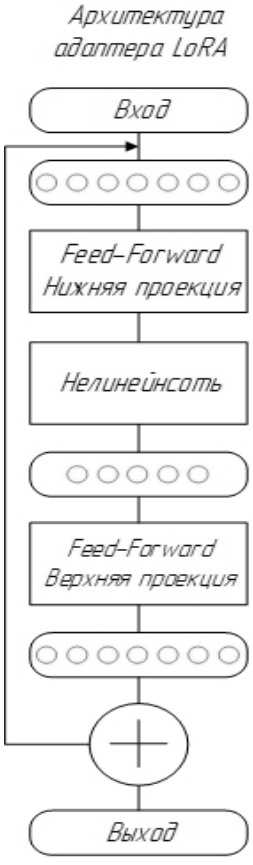

Рис. 1. Схемы архитектуры адаптера LoRA

Кроме того, применение LoRA позволяет улучшить качество генерируемых изображений. Низкоранговые слои, полученные с помощью LoRA, способны выделять более информативные признаки и улучшать качество генерируемых изображений.

Также, использование LoRA позволяет снизить риск переобучения модели, что может улучшить ее обобщающую способность и уменьшить ошибку на тестовых данных (Рис. 2).

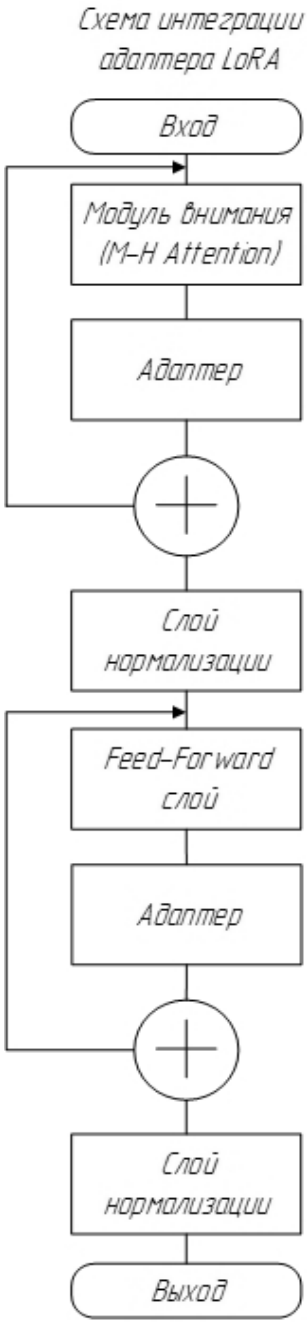

Рис. 2. Схемы интеграции адаптера LoRA

Математическая часть метода LoRA основана на принципе сжатия информации. Низкоранговые слои, полученные с помощью SVD или Input Compression Layers, позволяют сжать информацию о входных данных, сохраняя при этом их основные характеристики. Это позволяет уменьшить количество параметров модели, сохраняя ее способность к обучению (Рис. 3) [7].

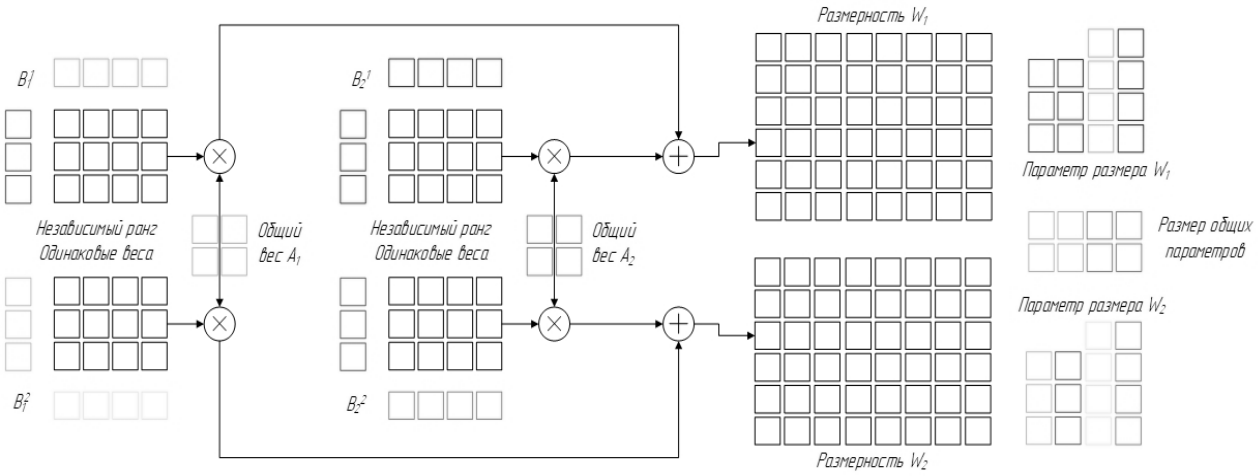

Рис. 3. Схема процесса генерации весов для двух слоев адаптера LoRA

Однако, существует дополнительный метод использования данной технологии – заранее создается и обучается самостоятельная модель LoRA, которая в последствии используется совместно с основной диффузионной моделью. Для этого необходимо сформировать небольшой датасет, состоящий из 16-24 пар типа изображение – текстовое описание, что означает необходимость сформировать текстовый файл с описанием изображенного объекта и поместить его рядом с самим изображением (Рис. 4).

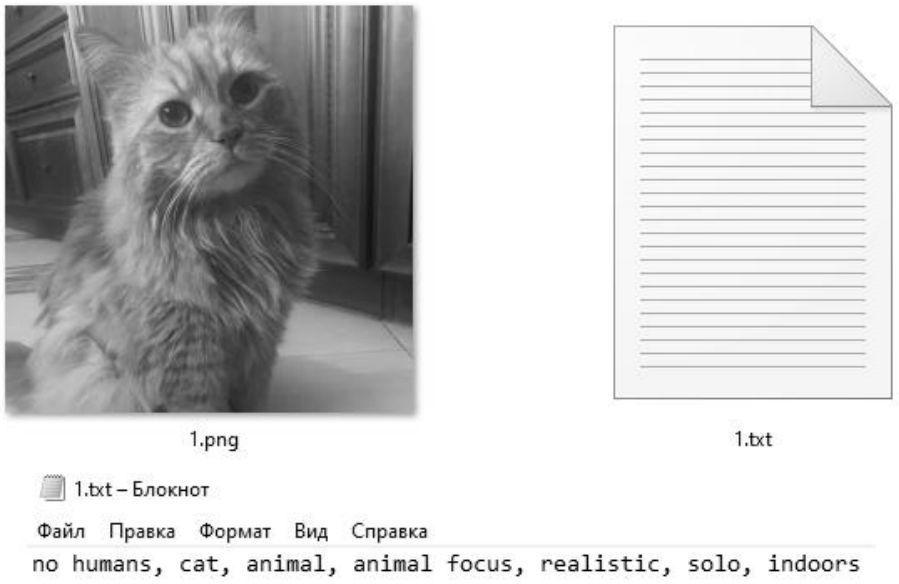

Рис. 4 Пример элемента обучающей выборки для модели с низкоранговой адаптацией.

В общем виде, принцип работы модели низкоранговой адаптации LoRA, совместно с диффузионной моделью выглядит следующим образом (Рис. 5).

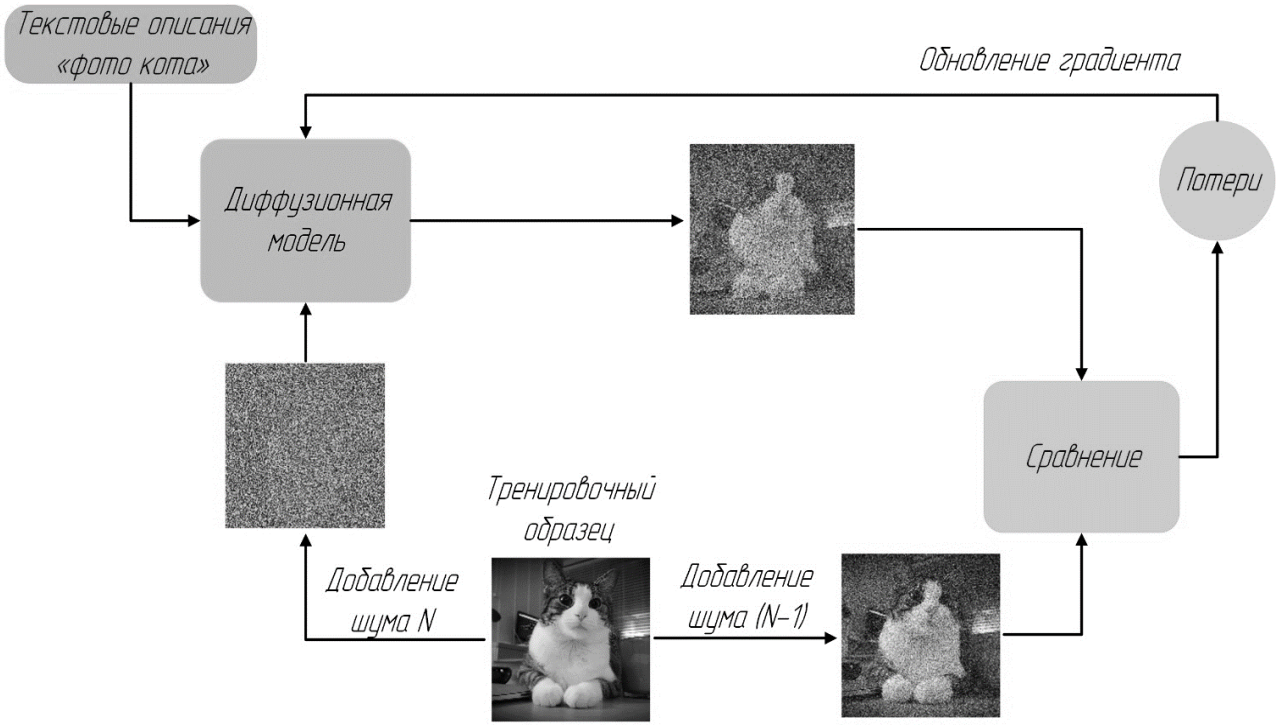

Рис. 5. Схема принципа работы LoRA с диффузионной моделью

Заключение: Применение низкоранговой адаптации LowRankAdaptation (LoRA) совместно с диффузионными моделями позволяет существенно снизить вычислительную сложность и ускорить процесс обучения. Низкоранговые слои, полученные с помощью LoRA, позволяют заменить полносвязные слои в архитектуре диффузионной модели, что позволяет снизить количество параметров и улучшить качество генерируемых изображений. Кроме того, использование LoRA позволяет снизить риск переобучения модели и улучшить ее обобщающую способность.

По мимо этого существует возможность создания модели LoRA, что позволяет направлять основную диффузионную модель на необходимые пользователю объекты, что в перспективе применимо для формирования новых датасетов определённых предметов и т.д. Благодаря своей высокой производительности и качеству генерируемых изображений, метод LoRA может быть полезным инструментом в таких областях, как компьютерное зрение, машинное обучение и искусственный интеллект.

Таким образом, использование низкоранговой адаптации LowRankAdaptation (LoRA) совместно с диффузионными моделями позволяет существенно улучшить процесс генерации изображений по текстовому описанию. Метод LoRA позволяет снизить вычислительную сложность, ускорить процесс обучения, улучшить качество генерируемых изображений и снизить риск переобучения модели.

Библиографический список:

1. Zhao Y., Li J., Gong Y. Low-rank plus diagonal adaptation for deep neural networks // 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Shanghai, China, 2016, pp. 5005-5009.

2. Левин А.О., Белов Ю.С. Использование генеративно-состязательных сетей для генерации изображений по тексту // Научное обозрение. Технические науки. 2023. № 2. С. 11-15.

3. Liu X., More Control for Free! Image Synthesis with Semantic Diffusion

Guidance // 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023, pp. 289-299.

4. Дроздов Д.С., Белов Ю.С. Обзор подходов к построению моделей для генерации изображений по текстовому описанию // В сборнике: Фундаментальные и прикладные исследования. Актуальные проблемы и достижения. сборник статей всероссийской научной конференции. Санкт-Петербург, 2023. С. 28-30.

5. Gu S. Vector Quantized Diffusion Model for Text-to-Image Synthesis // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10686-10696.

6. Zhang C., Peng Y. Stacking VAE and GAN for Context-aware Text-to-Image Generation // 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), Xi'an, China, 2018, pp. 1-5.

7. Rombach R., Blattmann A., Lorenz D. High-Resolution Image Synthesis with Latent Diffusion Models // 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 2022, pp. 10674-10685.