УДК 004.8

ИССЛЕДОВАНИЕ ДООБУЧЕНИЯ ДИФФУЗИОННЫХ МОДЕЛЕЙ НА МЕДИЦИНСКИЙ ДОМЕН

А.Е. Бурлаков[1]

[1]*Самарский университет, г. Самара*
e-mail: Alexey.Burlakoww@gmail.com

Аннотация

Работа посвящена дообучению предварительно обученных диффузионных моделей для генерации на домен медицинских изображений с патологиями сетчатки. Проводится анализ выбранных открытых наборов данных, взятых для тонкой настройки, архитектурных отличий рассматриваемых моделей, процесс тонкой настройки моделей и оценка качества сгенерированных изображений с помощью LDA, Precision and Recall Distribution и фундаментальной модели RETFound.

Ключевые слова: диффузионные модели, тонкая настройка, сетчатка, LoRA адаптер, искусственный интеллект.

Введение

Офтальмологические заболевания существенно снижают качество жизни человека. Использование глубокого обучения позволяет автоматизировать диагностику по изображениям глазного дна (CFP), однако разработка моделей ограничена нехваткой размеченных данных и юридическими аспектами конфиденциальности. Синтезирование высококачественных изображений с помощью диффузионных моделей является перспективным подходом к решению этой проблемы. Целью данной работы является исследование эффективности адаптации диффузионных моделей к медицинскому домену с использованием LoRA-адаптера для генерации ранее неизвестных системе изображений.

1. Анализ данных

В качестве обучающего набора использован DeepEyeNet, а для тестирования – ODiR. Набор DeepEyeNet, предложенный в статье DeepOpht [1], содержит, 13 898 цветных изображений глазного дна (Color Fundus Photography, CFP) и 1811 изображений флюоресцентная ангиография (Fluorescein Angiography, FA) для 265 уникальных патологий сетчатки, аннотированные профессиональными заключениями опытных офтальмологов. Набор ODiR был представлен в рамках Peking University International Competition on Ocular Disease Intelligent Recognition (ODIR-2019) [2] и содержит 5 814 CFP, разделённый на 7 патологий и 1 класс нормального глазного дна. Для использования обучающего набора данных в задаче тонкой настройки моделей и последующей оценки сгенерированных изображений выполняется предварительная обработка данных. Предварительно обработанный набор данных включает в себя 306, который содержит 104 изображений и текстовых описаний с патологий возрастной макулярной дегенерацией, 177 изображений и текстовых описаний с патологией диабетической ретинопатии и 25 изображений и описаний с патологией миопии.

2. Выбор моделей и метод дообучения

В рамках данного исследования были отобраны три архитектурные модели, представляющие различные технологические подходы к генерации: Stable Diffusion 1.4, Kandinsky 2.2 и FLUX.1-dev. Модель Stable Diffusion 1.4 выступает в качестве классической архитектуры скрытой диффузии (Latent Diffusion Model), переносящей вычислительные операции в сжатое пространство. Kandinsky 2.2, характеризуется наличием промежуточного этапа Image Prior, в рамках которого текстовый эмбеддинг проецируется в визуальное пространство CLIP перед началом диффузионного процесса. Модель FLUX.1-dev отличается интеграцией диффузионного трансформера взамен традиционной архитектуры U-Net и использованием механизма Flow Matching, направленного на аппроксимацию оптимальной траектории перехода от шума к распределению данных. Для тонкой настройки выбранных моделей на специализированный набор данных изображений сетчатки использовался метод LoRA (Low-Rank Adaptation) [3]. Данный метод добавляет матрицы к весам моделей, основные веса моделей замораживаются, обучается только LoRA.

3. Анализ результатов

В данной работе для извлечения признаков используется модель RETFound [4]. Это фундаментальная модель, обученная на не аннотированных 904 170 CFP и 736 442 изображений оптической когерентной томографии глазного дна. Извлечённые признаки с тестового и сгенерированного набора классифицируются при помощи линейного дискриминантного анализа (LDA). Как видно из рисунка 1, модель Stable Diffusion 1.4 меньше всего смогла соответствовать распределению признаков, когда как модель FLUX.1-dev лучше всего покрыла тестовое распределение признаков. На рисунке 2 отображается визуальное сравнение генерации моделей до их тонкой настройки и после. Для получения всех изображений использовалось одно текстовое описание: «Diabetic retinopathy, mild, microaneurysms, retinal hemorrhages, venous beading, retinal edema.».

Для дополнительного анализа сгенерированных изображений используется метрика Precision and Recall Distribution (PRD) [5]. Извлечённые признаки из модели RETFound формируют множества реальных и сгенерированных векторов. Вычисляется попарное евклидово расстояние между всеми векторами множеств и для каждого вектора признаков формируется гиперсфера с радиусом k-го соседа. Авторы статьи рекомендуют выбор соседей равным 3, поэтому для оценивания берётся такое значение параметра. Метрики определяются по формулам (1). Результаты метрик приведены в таблице 1.

$$Precision(\Phi_r, \Phi_g) = \frac{\sum_{\varphi_g \in \Phi_g} f(\varphi_g, \Phi_r)}{|\Phi_g|} \quad Recall(\Phi_r, \Phi_g) = \frac{\sum_{\varphi_r \in \Phi_r} f(\varphi_r, \Phi_g)}{|\Phi_r|} \quad (1)$$

Precision характеризует долю сгенерированных изображений, попадающих в многообразие реальных данных, отражая правдоподобие и клиническую достоверность синтетических образцов. Recall оценивает, какая часть распределения реальных изображений покрыта генеративной моделью, характеризуя разнообразие синтеза и способность модели воспроизводить редкие варианты патологии. Для задачи синтетического расширения медицинских выборок критично высокое значение обеих метрик одновременно: низкая Precision означает риск появления артефактных изображений, не соответствующих реальной клинической картине, что способно исказить обучение последующих диагностических моделей, тогда как низкий Recall

указывает на схлопывание распределения и недостаточное покрытие вариативности патологии, что ограничивает практическую пользу дополненного набора.

С этой точки зрения полученные результаты демонстрируют существенные различия между рассмотренными архитектурами. Stable Diffusion 1.4 показывает наиболее слабые значения по обоим критериям, что согласуется с визуальной оценкой на рисунке 1 и свидетельствует об ограниченной применимости данной модели для расширения медицинских выборок. Kandinsky 2.2 обеспечивает заметно лучшее покрытие распределения за счёт этапа Image Prior, однако уступает в правдоподобии, что может быть связано с проекцией текстовых эмбеддингов в визуальное пространство CLIP, изначально ориентированное на общедоменные изображения. Модель FLUX.1-dev демонстрирует значимое превосходство по обеим метрикам, что объясняется использованием диффузионного трансформера и механизма Flow Matching.

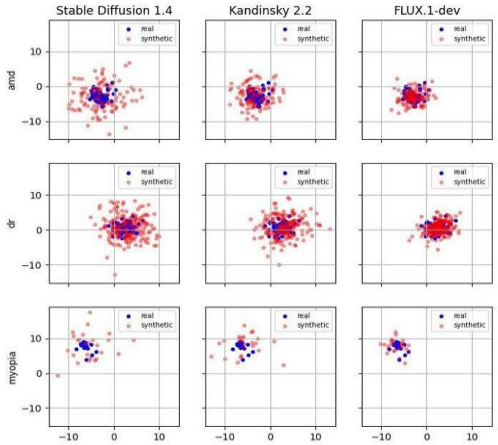


Рисунок 1 – Распределение классов сгенерированных и реальных изображений

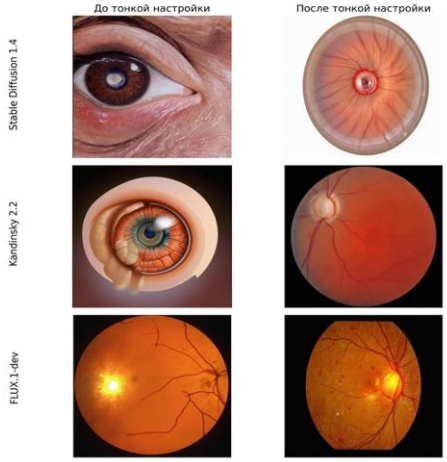


Рисунок 2 – Сравнение генерации до и после тонкой настройки

Таблица 1 – результаты метрик

Модель	**Precision**	**Recall**
Stable Diffusion 1.4	0,6734	0,7047
Kandinsky 2.2	0,7051	0,7925
FLUX.1-dev	0,8469	0,8763

Заключение

Полученные результаты показали, что различные архитектурные решения по-разному влияют на способность моделей воспроизводить распределение признаков реальных медицинских изображений. В частности, модель FLUX.1-dev продемонстрировала наилучшее соответствие распределению тестовых данных, тогда как Stable Diffusion 1.4 показала более выраженное отклонение в признаковом пространстве. Полученные значения PRD позволяют сделать практический вывод о применимости рассмотренных моделей для аугментации медицинских выборок: архитектура FLUX.1-dev обеспечивает наилучший компромисс между клинической достоверностью и разнообразием синтетических образцов, что делает её предпочтительным выбором для расширения обучающих данных диагностических систем при ограниченном объёме реальных изображений.

Список использованных источников

1. DeepOpht: Medical Report Generation for Retinal Images via Deep Models and Visual Explanation / J.-H. Huang, C.-H. H. Yang, F. Liu [et al.] // Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). – 2021. – P. 2442–2452.

2. ODIR-2019: Grand Challenge Dataset. – URL: https://odir2019.grand-challenge.org/dataset/ (дата обращения: 15.11.2025).

3. LoRA: Low-rank adaptation of large language models / E. J. Hu, Y. Shen, P. Wallis [et al.] // arXiv preprint. – 2021. – arXiv:2106.09685.

4. A foundation model for generalizable disease detection from retinal images / Y. Zhou, M. A. Chia, S. K. Wagner [et al.] // Nature. – 2023. – V. 622. – P. 156–163.

5. Improved Precision and Recall Metric for Assessing Generative Models / T. Kynkäänniemi, T. Karras, S. Laine [et al.] // NeurIPS. – 2019.

INVESTIGATION OF DIFFUSION MODEL FINE-TUNING FOR THE MEDICAL DOMAIN

A.E. Burlakov[1]

[1]*Samara National Research University, Samara*
e-mail: Alexey.Burlakoww@gmail.com

Abstract

This study focuses on the fine-tuning of pre-trained diffusion models for image generation within the medical domain, specifically targeting retinal pathologies. The paper provides an analysis of selected open-source datasets used for adaptation, examines the architectural differences between the models under consideration, describes the fine-tuning process, and evaluates the quality of the generated images using LDA, Precision and Recall Distribution and the RETFound foundation model.

Keywords: Diffusion models, fine-tuning, retina, LoRA adapter, artificial intelligence.