

ПОСТРОЕНИЕ СИСТЕМЫ ПОИСКА ПОХОЖИХ ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ FAISS И CLIP: ЭФФЕКТИВНЫЙ И МАСШТАБИРУЕМЫЙ ПОИСК ИЗОБРАЖЕНИЙ С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ И ВЕКТОРНОГО ПОИСКА

Новичков Д.Е. (Университет ИТМО)

Научный руководитель – к.т.н. Евстафьев О.А.
(Университет ИТМО)

Введение. В эпоху цифровой трансформации объемы визуальных данных растут экспоненциально, что делает задачи поиска и анализа изображений критически важными для множества приложений, таких как электронная коммерция, медиа, медицина и искусственный интеллект. Традиционные методы поиска изображений, основанные на текстовых метаданных или ручных аннотациях [1], часто оказываются недостаточно эффективными, особенно когда требуется семантическое понимание содержимого изображений. В данной работе представлен подход к созданию системы поиска похожих изображений с использованием современных технологий: CLIP (англ. Contrastive Language–Image Pretraining) для генерации семантических векторных представлений и FAISS (англ. Facebook AI Similarity Search) для оптимального поиска ближайших соседей. Этот подход позволяет создавать масштабируемые, точные и гибкие системы поиска изображений.

Основная часть. Современные технологии компьютерного зрения и искусственного интеллекта предоставляют новые возможности для организации и поиска визуальной информации. Одним из наиболее перспективных направлений является использование методов семантического поиска, в котором сравнение изображений осуществляется не только на основе их визуального сходства, но и с учетом их смыслового содержания. В данной работе исследуется и разрабатывается система поиска похожих изображений, основанная на интеграции передовых методов глубокого обучения и высокоэффективных алгоритмов векторного поиска. В качестве основных инструментов выбраны модель CLIP (англ. Contrastive Language–Image Pretraining) [2] и библиотека FAISS (англ. Facebook AI Similarity Search) [3], которые обеспечивают результативное извлечение семантических признаков изображений и быстрый поиск ближайших соседей в многомерном пространстве.

В ходе исследования были рассмотрены и протестированы различные подходы к обработке изображений, построению их векторных представлений и организации поиска. В первую очередь, особое внимание было уделено методу извлечения признаков, поскольку от его качества зависит точность поиска. Модель CLIP, разработанная OpenAI (2021), была выбрана в качестве основной технологии для создания семантических векторных представлений изображений. Данный метод позволяет результативно сопоставлять изображения и текст в едином пространстве признаков, что значительно расширяет возможности поиска, включая мультимодальный поиск по описаниям. Для индексации векторных представлений была использована библиотека FAISS, специально разработанная для поиска по многомерным данным. В работе были протестированы несколько стратегий индексирования:

- точный поиск (англ. Flat) — метод полного перебора, обеспечивающий максимальную точность, но требующий значительных вычислительных ресурсов;
- инвертированные файловые индексы (англ. Inverted File Indexing, IVF) — компромиссный вариант, позволяющий балансировать между скоростью и точностью;
- иерархические малые миры (англ. Hierarchical Navigable Small World, HNSW) — структура данных, ускоряющая поиск за счет построения графа соседей, что делает ее оптимальной для работы с большими объемами информации.

Данные для экспериментов были собраны из нескольких источников, включая открытые наборы данных изображений и пользовательские коллекции, содержащие фотографии различных категорий: товары электронной коммерции, медицинские снимки, изображения из социальных сетей и прочие визуальные данные. Оптимизация и производительность системы Одним из ключевых аспектов разработки стала масштабируемость системы. Для достижения высокой скорости обработки и поиска были применены следующие методы оптимизации:

- использование графических процессоров (GPU) для ускоренного извлечения признаков и работы с FAISS [4];
- пакетная обработка данных для уменьшения задержек при индексации;
- оптимизация параметров FAISS, таких как количество кластеров для IVF и глубина графа в HNSW.

Эксперименты показали, что предложенная система способна оптимально обрабатывать миллионы изображений, обеспечивая быстрый поиск с минимальной задержкой и точные результаты с высокой степенью соответствия запросу.

Выводы. Анализ результатов продемонстрировал значительное преимущество подхода, основанного на комбинации CLIP и FAISS, по сравнению с традиционными методами поиска изображений, использующими текстовые метаданные или ручную аннотацию. Применение нейросетевых моделей для извлечения признаков позволило повысить точность поиска за счет глубокого семантического анализа изображений [5]. Разработанная система может быть использована в различных областях:

- электронная коммерция — улучшение поиска товаров по изображениям;
- медицинская диагностика — поиск схожих снимков в базе данных для поддержки принятия решений врачами;
- медиа и журналистика — автоматическое сопоставление изображений с описаниями и поиск по контенту.

Список использованных источников:

1. Radford, A., Kim, J. W., Hallacy, C., et al. (2021). Learning Transferable Visual Models From Natural Language Supervision. *arXiv preprint arXiv:2103.00020*.
2. OpenAI. (2021). CLIP: Connecting Text and Images.
3. Facebook Research. (2021). FAISS: A library for efficient similarity search.
4. Johnson, J., Douze, M., & Jégou, H. (2019). Billion-scale similarity search with GPUs. *IEEE Transactions on Big Data*, 7(3), 535-547.
5. Wang, X., Zhang, L., Liu, Y., et al. (2022). Revisiting Image Retrieval with Transformers: A Benchmark and New Model. *arXiv preprint arXiv:2208.01079*