

Умная галерея: Семантический поиск по изображениям

Сибиряков Михаил, 24ТИДИ

Зачем это нужно?

- Проблема: у каждого из нас в галерее множество фотографий и часто мы помним, что изображено на фотографии, но не можем её найти так как не помним дату съемки или имя файла
- Цель: создать инструмент, который по пользовательскому описанию на естественном языке (русском) позволяет находить релевантные фото из каталога пользователя

Набросок реализации

В первом приближении нужно:

- Распознавать что находится на изображении (повесить тэги на каждое фото)
- Искать по слову или фразе релевантные фото
- Очевидно, простой поиск очень плох (похожие по смыслу слова игнорируют)
- Будем использовать векторные представления слов (Word Embeddings) для поиска по смыслу, а не символам
- Часто люди путают слова или плохо помнят что было на фото, но используют синонимы

Технологический стек

Проект объединяет в себе сразу две сферы машинного обучения: computer vision и natural language processing

- CV: для определения тэгов для фотографий использовалась готовая модель ResNet50 (обученная на ImageNet)
- NLP: модель FastText для “понимания смысла” слов и работы с русским языком
- Интерфейс: библиотека Tkinter для создания десктопного приложения (для демонстрации работы в первом приближении)

Архитектура системы

- Индексация: проставляем теги с помощью ResNet50, переводим их на русский, с помощью fasttext преобразовываем русски теги в векторы
- Сохраняем индекс в отдельный файл
- Так же логируем теги и их вероятности от ResNet50 в .json
- Поиск: векторизируем запрос пользователя и среди тегов, которые точно есть в индексе, ищем лучшие совпадения. Лучшие = близкие по смыслу. Для этого используем косинусное сходство

Демонстрация работы

AI Photo Finder

автобус



000000566758.jpg
Sim: 1.00



000000088462.jpg
Sim: 1.00



000000196843.jpg
Sim: 0.93



000000387387.jpg
Sim: 0.90



000000269632.jpg
Sim: 0.83



000000234660.jpg
Sim: 0.74

AI Photo Finder

транспорт



00000006040.jpg
Sim: 0.60



000000376625.jpg
Sim: 0.59



000000514376.jpg
Sim: 0.59



000000566758.jpg
Sim: 0.58



000000088462.jpg
Sim: 0.58



000000194832.jpg
Sim: 0.54



000000000000.jpg
Sim: 0.54



000000211674.jpg
Sim: 0.54



000000025181.jpg

COCO 2017 Dataset

ФИШКИ И СЛОЖНОСТИ

- Проблема точности: Как отсеять случайные распознавания?
- Решение: Формула $\text{Final Score} = \text{Semantic Similarity} \times \text{Model Confidence}$
- Перевод тезов не всегда удачный
- 1000 тезов – это довольно мало (многие из них повторяются почти – разные виды кошек, слонов)

Отсутствие метрик

- Задача не формализована (что значит «релевантная выдача»?)
- Основной критерий: пользовательский опыт
- Варианты метрик:
 1. Точность на первых k результатах (сколько из первых 10 фото релевантны запросу?)
 2. Насколько высоко первый релевантный результат?
- Не ясно, как выбирать запросы для оценки
- Сложно автоматизировать – нужен человек для проверки

Метрики

•

Запрос	Точность на первых 10	Первый релевантный
лыжи	1	1
кошка	0.7	1
овца	0.9	1
река	0.5	2
ноутбук	1	1
пес	0.3	1
цапля	0.4	1
жираф	0	11

Спасибо за внимание!