

Распознавание достопримечательностей на фотографиях

Смирдин А.Н.

Научный руководитель: к.ф.-м.н. Д. Н. Москвин

Научный консультант: В. В. Ермаков, Одноклассники

Санкт-Петербургская школа физико-математических и компьютерных наук

НИУ ВШЭ – Санкт-Петербург

2020

- Одноклассники – вторая по популярности социальная сеть на территории России. Месячная аудитория составляет более 70 миллионов пользователей
- Фотографии с достопримечательностями важны для пользователей, они создают повышенную активность
- Таких фото 3 – 5% от общего числа, что составляет миллионы фотографий ежемесячно
- Вместе с частью фотографией загружается так же метаинформация, в том числе место съемки
- Если дополнительно выделять фотографии с достопримечательностями, указывая, что на них изображено, то это дополнительно увеличит активность пользователей

Известные задачи:

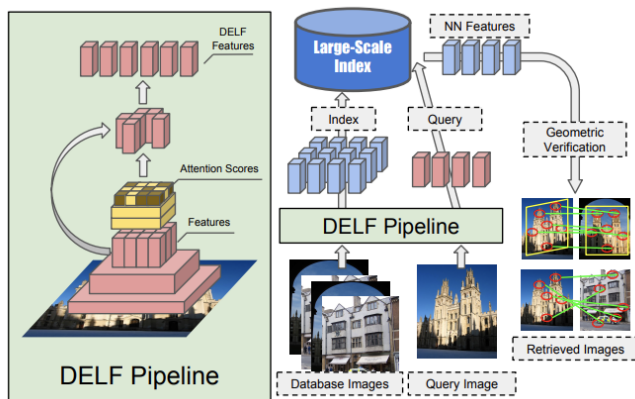
- Задача классификации изображений (Image Recognition)
- Близкой является задача поиска изображений похожих на данное (Image Retrieval)

Основные подходы:

- Можно решать с помощью техник компьютерного зрения
- Современные подходы основаны на глубоких сетях

Существующие решения: Deep Local Features (DELf)¹

- Глубокая сеть извлекает локальные признаки и их положения
- Поиск ближайших соседей по множеству признаков
- Геометрическая верификация



¹Hyeonwoo Noh et al. "Large-Scale Image Retrieval with Attentive Deep Local Features". In: *2017 IEEE International Conference on Computer Vision (ICCV)* (Oct. 2017).

Существующие решения: решение от Mail.ru²

- Одна сеть для выделения признаков и поиск ближайших соседей на множестве классов
- По векторным представлениям фотографий строится центроид для каждого класса
- Фотографии назначается класс ближайшего центроида
- Если до всех центроидов расстояние велико, то это не достопримечательность

²Andrei Boiarov and Eduard Tyantov. “Large Scale Landmark Recognition via Deep Metric Learning”. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management - CIKM '19* (2019).

Существующие решения: решения с Kaggle³

- Описаны способы чистки датасета
- Разные команды использовали разные архитектуры сетей и функции ошибки
- Отдельные сети, чтобы определять есть ли на фотографии достопримечательность
- Решения задачи поиска, а уже на ее основе задачи распознавания

³<https://www.kaggle.com/c/landmark-recognition-2019>

Требования к реализации

- Время обработки одной фотографии не должно превышать 0.3 с
- Классификация должна выполняться без использования GPS данных, но при наличии, они должны повышать точность
- Мало ложноположительных срабатываний

Ограничения существующих решений:

- DELF и решения с Kaggle имеют длительное время работы
- Решение и датасет Mail.ru не находятся в открытом доступе

Цель: создать модель для распознавания достопримечательностей, которая будет удовлетворять перечисленным требованиям

Задачи:

- Создать размеченный датасет с достопримечательностями
- Выбрать архитектуру глубокой сети и обучить модель
- Сравнить результаты с существующими решениями

Создание датасета: уже существующие данные

Датасет	#Изображений	#Объектов
Google-Landmarks-v1	1,225,029	14,951
Google-Landmarks-v2	4,132,914	203,094
Google-Landmarks-v2 очищенный	1,920,676	104,912
Yandex	213,678	672
Oxford5k	5062	11
Paris6k	6412	12

- Выбран уже очищенный датасет, в котором удалено большинство нерелевантных фотографий: планы зданий, экспонаты музеев
- Изображения в датасете взяты со страниц Wikimedia, для 89% достопримечательностей есть информация об их координатах
- Удалены кладбища, сады, парки, леса, реки

Создание датасета: данные для тестирования

- Датасет для тестирования надо построить из отобранных по гео-данным фотографий пользователей Одноклассников
- Среди них надо выбрать те, на которых действительно изображены достопримечательности
- Для достижения максимальной точности решено использовать ручную разметку
- Вручную отобрано по одному представителю для 700 классов
- Ассесорам показывается фотография эталон и фотография, которую надо разметить. Человек должен определить изображен ли на них один и тот же объект
- Из 20000 фотографий 5000 были определены как содержащие достопримечательности

Создание датасета: фотографии пользователей

- Результаты на тестовом датасете хуже чем на Google-Landmarks. Нужно дополнять датасет фотографиями из Одноклассников
- Фотографий пользователей много: в среднем по 2000 фотографий для каждой достопримечательности
- Данные последовательно фильтруются в два этапа:
 - Моя модель: быстрая, но не очень точная фильтрация
 - DELF: медленная и точная фильтрация

Итоговый датасет для обучения:

- 1983 достопримечательности из России и Европы
- 279232 фотографии достопримечательностей
- 227339 фотографий без достопримечательностей

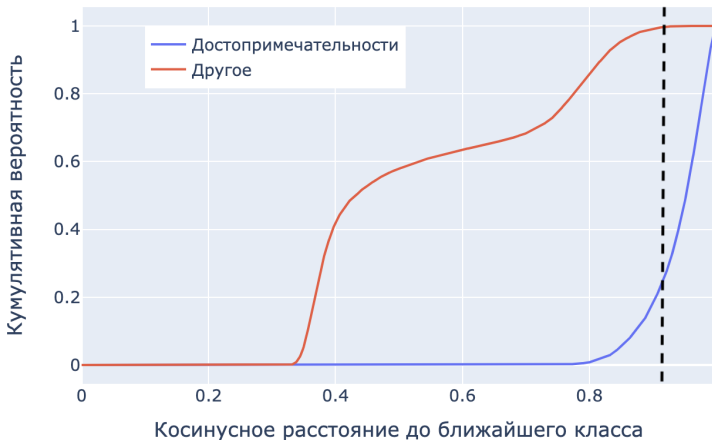
Тестовый датасет:

- 4762 фотографии с достопримечательностями

- Решение основано на подходе, предложенном Mail.ru
- ResNet-50 предобученная на датасете ImageNet
- Разные функции ошибки: Cross Entropy Loss, Center Loss, ArcFace Loss, CosFace Loss
- Определение класса, исходя из расстояний до центроидов
- Как определить изображена ли достопримечательность на фото?

Функция потерь	Точность
Cross Entropy Loss	82.4%
Center Loss	87.6%
ArcFace Loss	85.1%
CosFace Loss	84.7%

Достопримечательность или нет?



На валидационных данных подбирается такое минимальное расстояние до ближайшего класса, чтобы 99% не достопримечательностей классифицировались верно.

- DELF самая цитируемая работа в этой области
- Никак не надо переучивать модель, извлекающую признаки, под построенный датасет. Надо только извлечь признаки из всех фотографий

Модель	Точность	Полнота	Обработка одной фотографии, с
DELF	94.1%	75.5%	0.70
Эта работа	95.3%	72.7%	0.23

- Разработан автоматический способ создания датасета. Построенный датасет содержит более чем в 3 раза больше примеров чем соответствующие классы в Google-Landmarks
- Созданная модель удовлетворяет ограничению по времени в 0.3с на обработку одной фотографии
- Модель сравнялась по точности с одним из лучших решений в области, но работает быстрее (0.23с вместо 0.7с)
- Разработанное решение планируется к внедрению в социальной сети Одноклассники

Репозиторий: github.com/AndreySmirdin/LandmarkRecognition

Модель	Данные	Точность
DELF	Google-Landmark	82.6%
DELF	Google-Landmark + OK	85.1%
Эта работа	Google-Landmark	78.2%
Эта работа	Google-Landmark + OK	86.0%

Сравнение архитектур сети

