

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский университет
«Высшая школа экономики»

Факультет компьютерных наук
Основная образовательная программа
Прикладная математика и информатика

ГРУППОВАЯ КУРСОВАЯ РАБОТА

ПРОГРАММНЫЙ ПРОЕКТ НА ТЕМУ

**"РЕШЕНИЯ НА ОСНОВЕ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ ПРОЕКТОВ В
ГОРОДСКОЙ СРЕДЕ"**

Выполнили студент группы 171, 3 курса,
Биршерт Алексей Дмитриевич,
Шабалин Александр Михайлович

Руководитель КР:
старший преподаватель Соколов Евгений Андреевич

Консультант:
научный сотрудник Лобачева Екатерина Максимовна

Москва 2020

Содержание

1	Введение	3
2	Обзор литературы Алексей Биршерт (TODO я хз надо у Кати спросить)	4
3	Основная часть Алексей Биршерт (TODO я хз надо у Кати спросить)	4
3.1	Описание метода	4
3.2	Подготовка данных	5
3.3	Обучение моделей	5
3.4	Обработка результатов	6
3.5	Выводы	6

Аннотация

При проведении антропологических исследований часто необходимо обрабатывать данные внушительных объемов. Существует вариант обработать все эти данные вручную, однако такой метод слишком затратен по времени и трудовым ресурсам. В нашей проектной работе мы решаем две задачи автоматизации обработки данных. Первая - проверка того, что выделенное на фотографии лицо принадлежит живому человеку, а не напечатано на баннере или является скульптурой. Решение этой задачи позволяет уменьшить шум в данных и повысить точность статистик возрастных и гендерных групп, вычисляемых по фотографиям. Вторая - классификация достопримечательностей по фотографии. Решение этой задачи, в свою очередь, позволяет автоматизировать подсчёт статистик популярности публичных мест. В своих решениях мы используем сверточные нейронные сети различных архитектур.

Ключевые слова—Распознавание лиц, Глубокое обучение, Антропология

When conducting anthropological studies, it is often necessary to process a large amount of data. There is an option to process all this data manually, but this method is too time-consuming and labor-intensive. In our project, we solve two problems of data processing automation. The first is to verify that the face highlighted in the photo belongs to a living person, and is not printed on a banner or is a sculpture. The solution to this problem allows us to reduce noise in the data and improve the accuracy of gender and age groups statistics calculated from photographs. The second is the classification of sights in the photograph. The solution to this problem, in turn, allows you to automate the calculation of statistics on the popularity of public places. In our solutions we use convolutional neural networks of various architectures.

Keywords—Face Recognition, Deep Learning, Anthropology

1 Введение

На сегодняшний день человечество владеет огромными объемами данных, и во многих сферах деятельности приходится каким-либо образом с ними взаимодействовать. Одной из таких сфер является антропология. Для изучения человеческого развития и культуры необходимо наблюдать за человеком и анализировать его поступки и предпочтения. Одной из задач антропологов является облагораживание города. Для ее выполнения им нужно знать, где горожанам не хватает детской площадки или парка, где требуется произвести ремонт или реконструкцию здания. Хорошим источником информации о людях являются фотографии жителей какого-либо населенного пункта, ведь из них можно узнать, какие достопримечательности или объекты архитектуры наиболее привлекают людей, к каким возрастным и гендерным группам относятся эти люди, а также еще много другой важной для антропологии статистики. Конечно, никто не хочет тратить свое время и силы на просмотр тысяч фотографий и выделение из них полезных данных, гораздо удобнее и выгоднее автоматизировать этот рутинный процесс там, где это возможно. Существует довольно много известных методов выделения лиц на фотографиях. Однако, такие подходы выделяют также лица памятников и лица, напечатанные на рекламных щитах и вывесках. В нашей работе мы хотим отсеивать такие ошибки. Таким образом, первая задача нашего проекта - получение алгоритма, способного отличить “живое” лицо от “неживого”. Вторая задача – выделение и классификация архитектурных сооружений на фотографиях. Решение данной задачи позволит автоматически выделять множество фотографий, содержащих конкретный интересующий объект для проведения дальнейшего анализа. В ходе выполнения работы мы не будем предлагать каких-либо новых методов, однако мы используем ряд различных уже существующих технологий машинного обучения, которые могут быть применимы в решении самых разных задач. Например, ... В итоге мы получим алгоритм, способный перебирать большие объемы фотографий,

находить на них нужные объекты и собирать важные статистики для помощи в проведении исследований.

2 Обзор литературы Алексей Биршерт (TODO я хз надо у Кати спросить)

На данный момент в области распознавания лиц применяются сверточные нейросети. Так как моя задача связана с классификацией лиц на живые и неживые (памятники и скульптуры), то я могу лишь позаимствовать некоторые подходы в выборе архитектур и методов обучения. В частности от статьи [Primate Face Identification in the Wild \(2019\)](#) было почерпнуто знание про модификацию архитектур нейросетей и идея про построение функции ошибки для увеличения количества отличительных признаков классов. На основе статьи [Deep Face Recognition: A Survey \(2018\)](#) была построена модель принятия решения по конкретному лицу и модель обучения нейросети. Так же на основе статьи [One Millisecond Face Alignment with an Ensemble of Regression Trees \(2014\)](#) была построена модель для выравнивания лиц.

3 Основная часть Алексей Биршерт (TODO я хз надо у Кати спросить)

3.1 Описание метода

Для принятия решения по конкретному изображению лица является ли оно человеческим или принадлежит скульптуре в своей работе я использую сверточные нейросети. В качестве основных изучаемых нейросетей я выбрал три архитектуры - ResNet-18, Densenet-121 и Densenet-201. В каждой модели был убран первый слой maxpool и последний полносвязный слой заменён на слой с двумя выходными нейронами. TODO дописать номально.

3.2 Подготовка данных

Во время работы я столкнулся с проблемой недостаточного количества данных в классе памятников - поиск баз данных с фотографиями скульптур людей в открытом доступе не дал положительных результатов. Было принято решение выкачать фотографии из тематических групп в социальной сети Вконтакте посвященных скульптуре и живописи. Таким образом набралось около 7000 различных фотографий скульптур и бюстов. В качестве датасетов с фотографиями людей было решено взять датасет Labeled Faces in the Wild - в нём было около 13 тысяч фотографий различных людей и датасет IMDB-WIKI, в нём было более 500 тысяч фотографий людей. Из каждого датасета я случайно отобрал по 10 тысяч фотографий для дальнейшей обработки.

Для выделения лиц из фотографий использовался детектор лиц на основе сверточной сети и модель для предсказания точек на лице на основе ансамбля решающих деревьев из библиотеки dlib. Для обучения последней был использован датасет iBug 300-W.

После выделения области с лицом с помощью детектора я находил положение 12 ключевых для меня точек на лице с помощью [модели \(2014\)](#) - по 6 на каждый глаз. После я производил выравнивание лица, чтобы прямая проведенная через центры глаз была горизонтальна и левый глаз находился на определенном расстоянии от края фотографии, для баланса размера лиц. После полученное изображение сохранялось на диск для дальнейшего использования. После обработки всех фотографий и ручной очистки данных от мусора было получено следующее количество образцов классов - около 2000 памятников и скульптур и чуть больше 20000 людей.

3.3 Обучение моделей

В своей работе я обучил три нейросети для бинарной классификации памятников и людей и одну модель на основе ансамбля решающих деревьев для предсказания положения глаз на фотографии. Для обучения нейросетей были

использованы обработанные и обрезанные ранее фотографии людей и памятников. Для увеличения обучающей выборки я использовал аугментацию по типу one-to-many, с помощью различных преобразований получающих новые объекты. В качестве преобразования каждый раз выбиралось одно из 4 следующих: случайное изменение яркости, контраста, насыщенности и оттенка картинки, случайное аффинное преобразование, случайное отражение вдоль вертикальной оси, проходящей через центр картинки, либо случайное преобразование в Grayscale. Потом все изображения приводились к размеру 100 на 100 пикселей и нормировались. Каждая нейросеть обучалась 64 эпохи с размером батча в 64 картинки и темпом обучения $1e-4$ с постепенным уменьшением коэффициента в 10 раз каждые 16 эпох. Модель для предсказания точек на лице обучалась с помощью встроенной в библиотеку dlib функции с параметрами по умолчанию.

3.4 Обработка результатов

В моем распоряжении имелся датасет с реальными данными, с которыми предстояло работать. Это были размеченные вручную фотографии людей. Таким образом я мог оценить точность своих алгоритмов в условиях поставленной задачи. TODO таблица + посчитать точность везде нормально + влияние выравнивания.

3.5 Выводы

Таким образом была решена задача бла бла бла с хорошим качеством. TODO

Список литературы

1. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Deep Residual Learning for Image Recognition. 2015
2. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. Densely Connected Convolutional Networks. 2016
3. Ankita Shukla, Gullal Singh Cheema, Saket Anand, Qamar Qureshi, Yadvendradev Jhala. Primate Face Identification in the Wild. In PRICAI 2019
4. Mei Wang, Weihong Deng. Deep Face Recognition: A Survey. 2018
5. Vahid Kazemi and Josephine Sullivan. One Millisecond Face Alignment with an Ensemble of Regression Trees. 2014