

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

"КЛАССИФИКАЦИЯ ПОЛА И ВОЗРАСТА ЛЮДЕЙ НА ФОТОГРАФИИ"

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

Руководитель КР: старший преподаватель
Соколов Евгений Андреевич

Куратор: Магистр, руководитель группы нейросетевых технологий
компьютерного зрения
Овчаренко Сергей Александрович

Москва 2020

Содержание

1	Введение	3
2	Обзор литературы	5
2.1	Классификация возраста и пола	5
3	Детектирование лица	6
4	Классификация гендерных и возрастных групп	7
4.1	Описание метода	7
4.2	Описание данных	7
4.3	Постановка задач обучения	8
4.4	Эксперименты	9
4.5	Результаты	9
5	Список литературы	10

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 задачу автоматизации обработки изображений. Наша задача заключается в классификации гендерных и возрастных групп людей, запечатленных на фотографии. В процессе решения возникли несколько подзадач. Первая - выделение лица на изображении. Решение этой задачи позволяет сконцентрироваться на самых важных для нас признаках. Вторая - проверка того, что выделенное на снимке лицо принадлежит живому человеку, а не напечатано на рекламном щите или является скульптурой. Решение этой задачи позволяет уменьшить шум в данных и повысить точность статистик возрастных и гендерных групп, вычисляемых по фотографиям. В своём решении мы комбинируем различные известные подходы для достижения наилучшего результата.

Ключевые слова—Определение возраста и пола, Распознавание лиц, Компьютерное зрение, 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 work, we solve the problem of automating image processing. Our task is to classify the gender and age groups of people captured in the photograph. In the process of solving several subtasks arose. The first is the selection of the face in the image. The solution to this problem allows us to concentrate on the most important features. The second is to verify that the face highlighted in the picture belongs to a living person, and is not printed on a billboard or is a sculpture. The solution to this problem allows us to reduce noise in the data and improve the accuracy of statistics of age and gender groups calculated from photographs. In our decision, we combine various well-known approaches to achieve the best result.

Keywords—Age and gender classification, Facial recognition, Computer vision, 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. Никто не хочет тратить свое время и силы на просмотр тысяч фотографий и выделение из них полезных данных, когда гораздо удобнее и выгоднее автоматизировать этот рутинный процесс там, где это возможно.

Все известные подходы к классификации возрастных групп людей по фотографии заключаются в анализе изображения лица. Самые ранние ([1]) основывались на различиях в пропорциях и размерах черт лица в зависимости от возраста - так называемые антропометрические модели ([2]). Все они вычисляли координаты точек на лице и в дальнейшем их анализировали. Более поздние методы ([3]) опирались на использование сверточных нейронных сетей различной глубины или полнокомпонентных сверточных нейронных сетей ([4]).

Первые подходы к классификации пола использовали фотографии лица низкого разрешения - до 20 на 20 пикселей, и обучали на них различные типы классификаторов ([5]). Позднее стали использовать LBP для выявления новых признаков ([6]), использовать сверточные нейронные сети ([3, 4]).

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. В итоге мы получим алгоритм, способный перебирать большие объемы фотографий, находить на них нужные объекты и собирать важные статистики для помощи в проведении исследований.

Дальнейшая работа описана в следующих главах - обзор литературы, распознавание лиц, отличие "живого" лица от напечатанного, классификация лиц людей и скульптур, классификация гендерных и возрастных групп. Отличие "живого" лица от напечатанного выполнено Александром Шабалиным, классификация лиц людей и скульптур Алексеем Биршертом.

2 Обзор литературы

2.1 Классификация возраста и пола

Задачи определения пола и возраста человека находят применение в разных сферах жизни человека, в наружном наблюдении, в антропологии, в биометрической идентификации. Современные подходы к этой задаче опираются на сверточные нейронные сети. Так, например, в статье [3] описано решение с помощью сверточной сети небольшой глубины. Для классификации возраста и пола используется одна и та же архитектура. Нейронная сеть состоит из трёх свёрточных слоёв и двух полносвязных, небольшой размер сети объясняется желанием быть физичным в распознавании лиц и нежеланием переобучиться. Точность по классификации пола была $86.8 \pm 1.4\%$, возрастных групп - $50.7 \pm 5.1\%$ для точного попадания в группу и $84.7 \pm 2.2\%$ для попадания в правильную или соседнюю.

В статье [4] описан алгоритм анализа лица с помощью пяти сверточных нейросетей, получающих изображения лица целиком, левого и правого глаза, носа и рта соответственно. Итоговое решение принимается на основе выходов всех пяти нейросетей. Нейросеть, получающая на вход всё изображение лица, имеет три сверточных слоя, прочие по два. Точность по классификации пола достигла $89.6 \pm 1.3\%$, возрастных групп - $54.3 \pm 3.5\%$ для точного попадания в группу и $87.6 \pm 1.9\%$ для попадания в правильную или соседнюю, что является значительным улучшением результата предыдущей статьи.

3 Детектирование лица

ЧТОТО СДЕЛАЛИ ДА

4 Классификация гендерных и возрастных групп

Эта часть выполнена Алексеем Биршертом.

4.1 Описание метода

В качестве базовой модели для распознавания пола и возраста используется глубокая нейронная сеть ResNet-18, последний полносвязный слой которой заменён на два полносвязных слоя с нелинейностью ReLU и дропаутом между ними. Последний слой содержит 2 выходных нейрона для классификации по полу (мужчина или женщина) и 101 для классификации по возрасту (от 0 до 100 лет включительно). На вход подаются фотографии лиц людей, детектированные с помощью модели распознавания лиц, размером 227 на 227 пикселей, 3 канала цвета - R, G, B. Предсказанный пол определяется с помощью определения выходного нейрона соответствующей нейронной сети с максимальным значением. Предсказанный возраст определяется с помощью преобразования вектора значений выходных нейронов соответствующей нейронной сети функцией софтмакс. Каждое значение вектора преобразуется в экспоненту в его степени, затем делится на сумму элементов преобразованного вектора. После вышеописанного преобразования мы получаем вектор вероятностей, которые модель выдала для каждого возраста в диапазоне от 0 до 100 включительно. Дальнейший подсчет точного предсказанного возраста получается с помощью подсчета матожидания возраста при соответствующем вероятностном распределении - каждый элемент вектора после применения софтмакса умножается на соответствующий ему возраст.

4.2 Описание данных

В качестве датасета для обучения двух вышеописанных моделей были избраны датасеты IMDB-WIKI-101 и FGNET. Распределение реального воз-

раста в датасете IMDB-WIKI-101 имеет вид нормальной кривой со средним около 35 лет, имея значительно малое количество объектов с возрастом меньше 10 лет или больше 90. Для восполнения данных по возрасту до 10 лет был избран датасет FGNET, в котором большая часть объектов это дети до 15 лет. Для улучшения сходимости нейронных сетей была произведена предобработка всех объектов - в итоговую выборку не были включены объекты следующие объекты: объекты с плохо различимыми лицами (показатель уверенности модели распознавания лиц в том, что это лицо ниже фиксированного значения), объекты с некорректно заполненными данными про пол/возраст, объекты со слишком маленькими фотографиями. Итого было получено около 200 тысяч объектов, которые были в дальнейшем поделены с сохранением баланса классов 1 к 19 на валидационную и обучающую выборки соответственно. В качестве датасета для тестирования был избран датасет Adience, по которому известно большое количество результатов различных моделей. Из него были исключены объекты с некорректным описанием пола или возраста. Итого было получено почти 11 тысяч объектов для тестовой выборки. В Adience метки возраста в формате 8 групп - 0: [0, 2], 1: [4, 6], 2: [8, 12], 3: [15, 20], 4: [25, 32], 5: [38, 43], 6: [48, 53], 7: [60, 100]. В связи с этим, необходимо было решить как относить метки реального возраста от 0 до 100 к этим группам. Было принято решение относить к ближайшей группе - например 22 года ближе к 20, чем к 25, следовательно группа 3. В случае одинакового расстояния выбиралась первая по порядку группа.

4.3 Постановка задач обучения

Задача классификации пола является задачей бинарной классификации, для обучения была выбрана перекрестная энтропия - функция ошибки со следующей формулой: $loss(x, y) = -x_y + \log \left(\sum_{j=1}^K x_j \right)$, где x - вектор значений выходных нейронов нейронной сети, y - правильный класс. Для оценки качества классификации использовались метрика доля правильных ответов

(далее accuracy), точность (далее precision) и полнота (далее recall).

Задача классификации возраста является задачей многоклассовой классификации. Если смотреть на задачу предсказания возраста по фотографии с точки зрения человека, человек гораздо точнее способен угадать диапазон возраста, нежели точный возраст. Поэтому задача классификации возраста была интерпретирована как задача с множественными правильными ответами - каждому объекту может соответствовать набор правильных классов. Для обучения была выбрана бинарная перекрестная энтропия - $loss(x, y) = \sum(L)$, $L = \{l_1, \dots, l_N\}$, $l_i = (y_i \log(x_i) + (1 - y_i) \log(1 - x_i))$, где x это вектор значений выходных нейронов нейронной сети, y - вектор нулей, содержащий единицы на местах правильных классов. Для оценки качества классификации использовались средний модуль отклонения (далее MAE) и доля объектов, у которых MAE не превышает фиксированного числа (далее CS-5).

4.4 Эксперименты

Первым необходимо было решить вопрос архитектуры базовой модели, было принято решение остановиться на ResNet-18. Далее нужно было решить вопрос с разделением количеством моделей - использовать одну модель с двумя выходами или две модели, каждая с одним выходом. Процесс подбора гиперпараметров: Процесс выбора аугментации, ну тут просто про то, какую взял, ибо я особо не экспериментировал Процесс выбора метода обучения возраста и так далее

4.5 Результаты

Сравнение на Adience с другими статьями, показать, что моя модель предсказывает хорошо, что плохо, графики

5 Список литературы

Список литературы

1. Age Classification from Facial Images. Young H. Kwon, Niels da Vitoria Lobo. In IEEE 1994
2. Age and Gender Estimation of Unfiltered Faces. Eran Eidinger, Roeen Enbar, Tal Hassner. In IEEE 2014
3. Age and Gender Classification using Convolutional Neural Networks. Gil Levi, Tal Hassner. In IEEE 2015
4. Age/gender classification with whole-component convolutional neural networks (WC-CNN). Chun-Ting Huang, Yueru Chen, Ruiyuan Lin, C.-C. Jay Kuo. In IEEE 2018
5. Learning Gender with Support Faces. Baback Moghaddam, Ming-Hsuan Yang. In IEEE 2002
6. Demographic Classification with Local Binary Patterns. Zhiguang Yang, Haizhou Ai. In ICB 2007
7. Robust Real-Time Face Detection. Paul Viola, Michael J. Jones. In IEEE 2003
8. Face detection without bells and whistles. Makrus Mathias, Rodrigo Beneson, Marco Pedersoli, Luis Van Gool. In ECCV 2014
9. Accurate Face Detection for High Performance. Faen Zhang, Xinyu Fan, Guo Ai, Jianfei Song, Yongqiang Qin, Jiahong Wu. In ArXiv 2019
10. Deep Residual Learning for Image Recognition. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. In IEEE 2015
11. One Millisecond Face Alignment with an Ensemble of Regression Trees. Vahid Kazemi and Josephine Sullivan. In IEEE 2014

12. Face Anti-Spoofing Based on Color Texture Analysis. Zinelabidine Boulkenafet, Jukka Komulainen, Abdenour Hadid. In IEEE 2015
13. Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision. Yaojie Liu, Amin Jourabloo, Xiaoming Liu. In IEEE/CVF 2018
14. Deep Tree Learning for Zero-shot Face Anti-Spoofing. Yaojie Liu, Joel Stehouwer, Amin Jourabloo, Xiaoming Liu. In IEEE/CVF 2019
15. Primate Face Identification in the Wild. Ankita Shukla, Gullal Singh Cheema, Saket Anand, Qamar Qureshi, Yadvendradev Jhala. In PRICAI 2019
16. Deep Learning Face Representation by Joint Identification-Verification. Yi Sun, Xiaogang Wang, Xiaoou Tang. In NIPS 2014
17. A Discriminative Feature Learning Approach for Deep Face Recognition. Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. In ECCV 2016
18. Densely Connected Convolutional Networks. Gao Huang, Zhuang Liu, Laurens van der Maaten, Kilian Q. Weinberger. In IEEE 2016
19. NUAA Photograph Imposter Database
20. Labeled Faces in the Wild Database
21. IMDB-WIKI Database
22. 300 Faces In-the-Wild Database