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

Руководитель КР: старший преподаватель
Соколов Евгений Андреевич

Куратор: Магистр, разработчик решений компьютерного зрения
Черномордик Григорий Петрович

Москва 2020

Содержание

Аннотация

При проведении антропологических исследований часто необходимо обработать 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 Детектирование лица

Детектирование лица на фотографии - одна из древнейших задач компьютерного зрения, возникшая в 1990-х годах. Первые хорошие результаты появились в 2004 году. Описанный в статье [7] метод находил лица с помощью признаков Хаара, используя каскад детекторов, обученных алгоритмом AdaBoost. В 2014 году в статье [8] был предложен метод, использующий Deformable Parts Model (DPM). Его идея заключается в нахождении зависимостей между подвижными частями. Например, лицо представляется, как нечто, состоящее из глаз, носа, рта, расположенных в некотором антропоморфическом виде. Однако все описанные методы показывали довольно плохие результаты в сложных случаях, так как опирались на ограниченный, придуманный людьми набор признаков. Поэтому методы, основанные на сверточных нейронных сетях быстро вытеснили остальные. Лучшие известные на данный момент подходы описаны в статье [9]. Все из них используют нейронную сеть (обычно ResNet [10]) для получения признаков. Так как модель сама находит признаки и зависимости, результат получается лучше.

Немаловажной задачей является и выравнивание лиц. Самый быстрый метод получить выровненное лицо - определить ключевые точки на лице и преобразовать изображение чтобы эти точки были на заранее определенных местах. В статье [11] описан метод вычисления координат основных точек на лице человека - окаймляющих лицо, глаза, нос, рот и брови. Вычисление точек происходит с помощью каскада регрессоров, обучаемых с помощью градиентного бустинга.

2.2 Отличие "живого" лица от напечатанного

Задача определения является ли лицо на фотографии напечатанным или реальным довольно популярна, так как используется в Face Anti-Spoofing си-

стемах, предназначенных для обнаружения попытки взлома при прохождении биометрической идентификации. Существует несколько различных подходов к решению этой задачи. В статье [12] используются LPB (Local binary patterns - Локальные бинарные шаблоны). В статье [13] используется решение с помощью нейронных сетей, бинарно классифицирующих лица на настоящие и напечатанные. Однако оно оказывается чуть хуже решения с LBP. Данная область постоянно развивается вместе с улучшениями методов взлома. В статье [14] был предложен алгоритм на основе иерархических нейронных сетей, который кластеризует фотографии вместо классификации и показывает феноменальные результаты.

2.3 Классификация лиц людей и скульптур

Проблема определения того, принадлежит лицо на фотографии человеку или скульптуре не является популярной задачей, однако существует литература, описывающая подходы к работе с идентификацией лиц. В частности, в статьях [15], [16] и [17] описаны три различных варианта построения функции ошибки и метода обучения сверточных нейронных сетей для наилучшей классификации лиц. Всего описано три функции ошибки - попарная ошибка (сумма перекрестной энтропии и дивергенции Кульбака-Лейблера), contrastive loss (сумма перекрестной энтропии и L2 нормы) и center loss (сумма softmax ошибки и center loss). Соответственно нейронные сети обучаются с целью разнесения разных классов как можно дальше друг от друга и приближения объектов внутри класса к центру.

Самыми популярными вариантами архитектур нейронных сетей для решения данной задачи являются ResNet ([10]) и DenseNet ([18]). Обе архитектуры позволили обучать более глубокие нейросети с непрерывным увеличением качества - глубокие нейронные сети предыдущих архитектур с какого-то момента не могли соревноваться в качестве с менее глубокими моделями как на обучающей выборке, так и на тестовой.

2.4 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

В настоящий момент мы пользуемся детектором dlib, планируется попробовать разные варианты и улучшить точность выделения лиц на изображениях.

4 Отличие "живого" лица от напечатанного

Эта часть сделана Александром Шабалиным.

4.1 Описание метода

Задача определения является ли лицо на фотографии напечатанным или реальным довольно популярна, так как используется в Face Anti-Spoofing системах, предназначенных для обнаружения попытки взлома при прохождении биометрической идентификации. Однако чаще всего при распознавании используется видео или инфракрасное изображение. В нашем случае мы владеем только фотографией объекта, что сильно усложняет задачу, ведь мы не можем, например, улавливать моргания или пульс, что позволит нам определить, является ли человек живым, как в случае с видео.

После проведения исследований, было решено использовать LBP (Local binary patterns - Локальные бинарные шаблоны), трансформировав изображение из RGB в YCbCr (яркость, синяя и красная цветоразностные компоненты). Сущность LBP метода заключается в построении вектора чисел. В простейшим варианте это делается следующим образом: для каждого пикселя берутся n соседних пикселей. Затем обходим их все по кругу. Для каждого пикселя сравниваем его значение с центральным, если значение больше, то ставим 1, иначе - 0. Получаем n-битное двоичное число. Проделав такую операцию со всеми пикселями, построим гистограмму распределения чисел. Сконкотенируем гистограммы для всех 3-х цветовых каналов и получим искомый вектор. Трансформировав таким образом все картинки, мы получим набор признаков, на основе которых обучим метод опорных векторов. Лучшим ядром для SVM оказалось полиномиальное ядро.

Интуиция данного метода заключается в том, что живое человеческое лицо отражает свет неравномерно, в отличие от бумажной фотографии. Поэтому средний вектор живого лица будет отличаться от вектора напечатанного. Важно заметить, что мы будем считать только те числа, у которых в дво-

ичной записи идут сначала единицы, а потом нули. Также будем считать их инвариантными к повороту.

Такой метод действительно оказался рабочим. Я обучал модель на датасете NUAA Photograph Imposter Database ([19]). Он содержит фотографии 15-ти людей. Для каждого человека есть около 400 реальных фотографий и 500 снимков фотографий. На этом датасете мне удалось получить точность классификации 99.8%, что является отличным показателем. Также данный алгоритм работает гораздо быстрее нейросетей и результат не зависит от поворота или смещения лица на картинке. Однако фотографии в датасете отличаются от фотографий, на которых предстоит использовать алгоритм. Выяснилось, что на реальных изображениях результаты гораздо хуже - точность около 70%, что нас совершенно не устраивает. Причина такого падения качества скорее всего заключается в качестве фотографий (В датасете NUAA Photograph Imposter Database фотографии больше и их разрешение выше). Так как данное решение оказалось неподходящим для поставленной задачи, от него пришлось отказаться и перейти к использованию нейронных сетей.

4.2 Выводы

Опробованный метод с использованием LBP не дал желаемых результатов, хотя и показал себя хорошо на датасете для анти-спуффинга. Из-за этого от него пришлось отказаться. В будущем планируется попробовать другие описанные выше методы решения поставленной задачи.

5 Классификация лиц людей и скульптур

Эта часть сделана Алексеем Биршертом.

5.1 Описание метода

Для принятия решения по конкретному изображению лица, является ли оно человеческим или принадлежит скульптуре, в своей работе я использую сверточные нейронные сети. В качестве основных изучаемых нейросетей я выбрал три архитектуры - ResNet-18 ([10]), DenseNet-121 и DenseNet-201 ([18]). В каждой модели был убран первый слой maxpool (это позволило работать с изображениями меньшего размера) и последний полносвязный слой заменён на слой с двумя выходными нейронами.

5.2 Подготовка данных

Во время работы я столкнулся с проблемой нехватки 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 ([20]) - в нём было около 13 тысяч фотографий различных людей и датасет IMDB-WIKI ([21]), 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 ([22]).

После выделения области с лицом с помощью детектора я находил поло-

жение 12 ключевых для меня точек на лице с помощью модели ([11]) - 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.

5.3 Обучение моделей

В своей работе я обучил три нейронных сети 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 афинное 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 втроенной в библиотеку dlib функции с параметрами по умолчанию.

5.4 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. В ближайшее время планируется оценить точность по этому датасету и более корректно оценить точность на обучающей выборке с помощью разбиения на 5-6 фолдов, изучить влияние различных факторов модели на итоговый результат.

5.5 Выводы

Таким образом задача была решена, осталось оценить качество.

6 Классификация гендерных и возрастных групп

В настоящий момент мы не приступили к выполнению данной части, планируется закончить выполнение до конца апреля.

7 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, Roee 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, Lus 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