Санкт-Петербургский государственный университет

**Кафедра компьютерных технологии и систем**

**Сугоняев Андрей Дмитриевич**

**Выпускная квалификационная работа бакалавра**

**Распознавание лиц с помощью нейронных сетей**

Направление 02.03.02

Фундаментальная информатика и информационные технологии

ООП "Программирование и информационные технологии"

Научный руководитель:

к.ф.-м.н., доцент

Погожев С. В.

Санкт-Петербург

2019

Содержание

Введение 3

Обзор существующих решений 6

Постановка задачи 10

Теоретическая часть 11

Сверточные нейронные сети 11

Пакетная нормализация и групповая нормализация 11

Архитектуры, основанные на регрессии-классификации 13

Практическая часть 17

Обучающая выборка 17

Предобработка 17

Аугментация данных 17

Архитектура сети 18

Алгоритм обучения 19

Реализация 19

Тестирование сети и анализ результатов 20

Заключение 22

Список литературы 23

**Введение**

Распознавание человека по лицу становится все более востребованным в век цифровых технологий и, с ростом рынка мобильных и десктопных устройств, а также ростом мощности этих устройств, набирает популярность как метод человеко-машинного взаимодействия.

Широкое распространение распознавание личности по фотографии получило в социальных сетях, где цифровая фотография является одним из самых распространённых способов самовыражения. Пользователи выкладывают десятки миллионов фотографий каждый день, и довольно часто на этих фотографиях присутствуют лица, в том числе лица пользователей социальной сети. Поэтому в социальных сетях, таких как ВКонакте и Facebook, очень востребованы алгоритмы распознавания лиц, которые используются для автоматического детектирования лиц, для идентификации пользователя по лицу, для автоматической отметки друзей пользователей на фотографиях.

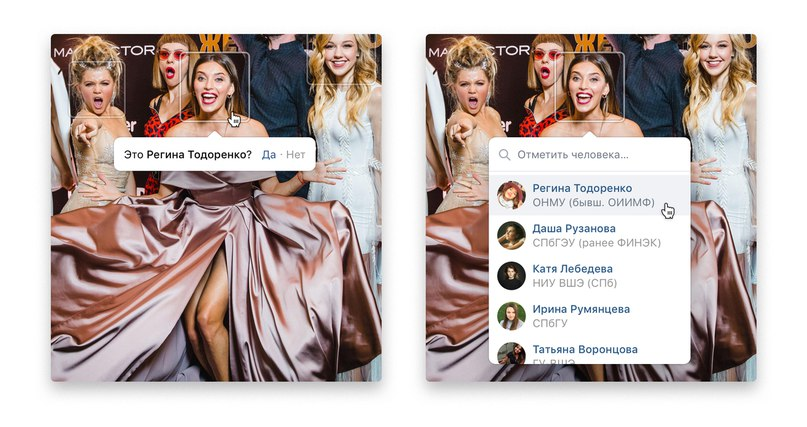


Рис. 1. Автоматическая отметка друзей на фотографии пользователя ВКонтакте

Помимо этого, распознавание лиц применяется в биометрических системах для подтверждения личности. Например, смартфоны Apple и Samsung используют распознавание лиц для разблокировки экрана.



Рис. 2. Работа системы Face ID компании Apple

Распознавание лиц является перспективной технологией в банковской сфере и сфере розничной торговли. Банки по всему миру внедряют технологии распознавания лиц для того, чтобы упростить процедуру оказания услуг и повысить безопасность. Многие магазины используют распознавание лиц для таргетирования рекламы, а некоторые используют терминалы с системой распознавания лиц для оплаты услуг.



Рис. 3. Терминал с системой распознавания лиц, расположенный на входе в один из китайских музеев

Распознавание лиц также применяется в решении задач правоохранительных органов. Использование алгоритмов автоматической идентификации личности в системах видеонаблюдения помогает в поиске преступников и в борьбе с терроризмом. Примером этого может служить внедрение алгоритмов распознавания лиц компании NtechLab в московскую городскую систему видеонаблюдения, эксплуатируемую Департаментом Информационных Технологий Москвы, в 2017 году [1].



Рис. 4. Работа системы распознавания лиц на кадре с камеры наблюдения в московском метро

**Обзор существующих решений**

Задачи идентификации лица и детектирования лиц существуют довольно давно и по отдельности хорошо изучены. Для решения этих задач предложено большое количество методов [2] [3]. Некоторые методы основаны на ручном извлечении признаков из фотографий, другие основаны на машинном обучении или глубинном обучении.

На сегодняшний день самым популярным методом решением задач распознавания лиц являются сверточные нейронные сети. Одно из главных преимуществ сверточных нейронных в том, что вся система распознавания может обучаться от начала до конца, от значений пикселей до конечных категорий. Главный их недостаток – это необходимость огромного числа размеченных данных для их обучения.

Использование сверточных нейронных сетей в задачах распознавания изображений берет начало в 1990-х. Сверточные сети со скользящим окном с переменным успехом применялись в задаче распознавания лиц в работах [4], [5], [6].

Широкое распространение глубокие сверточные сети получили после того, как AlexNet [7] продемонстрировала их высокую эффективность на большом наборе данных ImageNet в 2012 году. Эта сеть заняла 1 место в соревновании ImageNet Large Scale Visual Recognition Challenge 2012 как в секции классификации, так и в секции локализации.

На сегодняшний день типичная последовательность шагов, выполняемых системой распознавания лиц в процессе работы, выглядит следующим образом:

1. Детектирование лица (face detection)
2. Вырезание и выравнивание лица (face alignment)
3. Извлечение сверточных признаков (feature extraction)
4. Верификация/идентификация личности (classification)

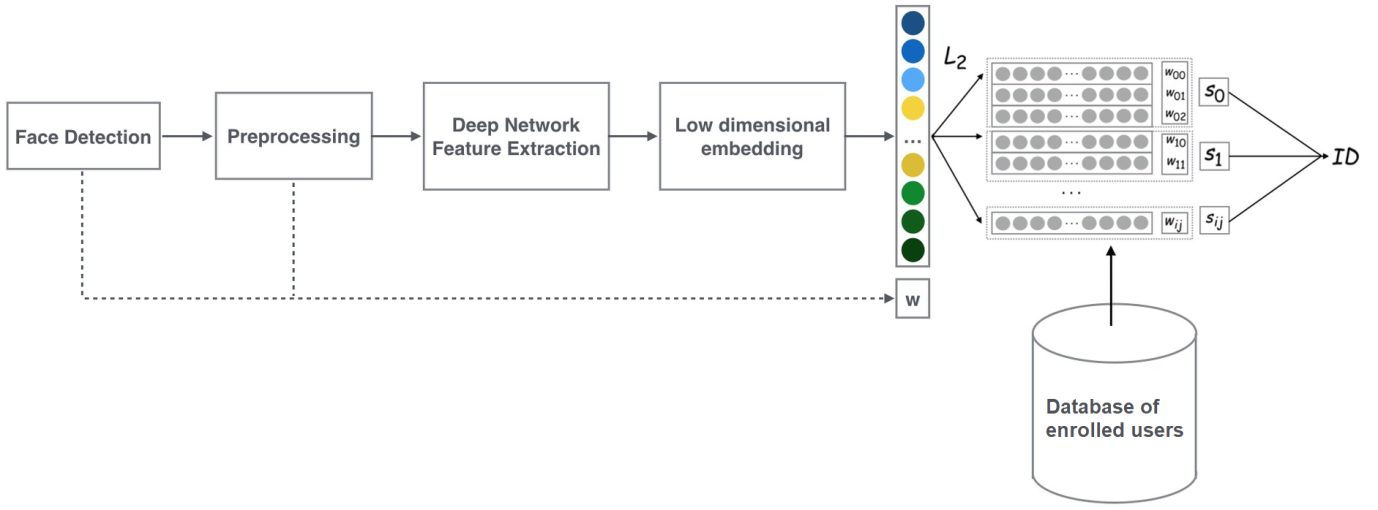


Рис. 5. Схема работы типичной системы распознавания лиц

Рассмотрим некоторые примеры систем распознавания лиц, применяемых крупными компаниями.

Система DeepFace [8], разработанная компанией Facebook в 2015 году, использует библиотеку dlib для детектирования лица и нахождения 67 ключевых точек. По этим точкам осуществляется трехмерное выравнивание. Затем система использует сверточную нейронную сеть VGG-16 для классификации.

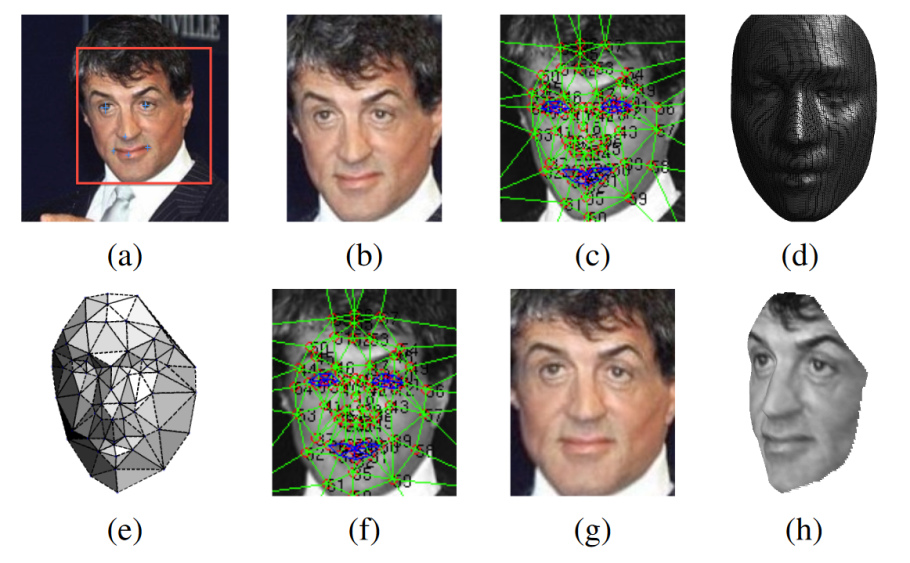


Рис. 6. Процесс трехмерного выравнивания лица по ключевым точкам

В “умных фотографиях” социальной сети ВКонтакте [9] для детектирования и выравнивания лиц используется каскад сверточных нейронных сетей Multi-Task CNN, а для классификации используется сверточная сеть ResNet.

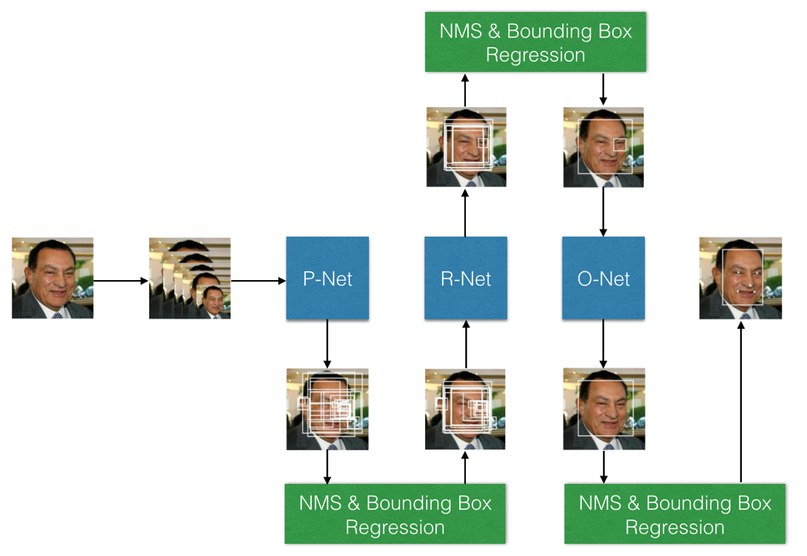


Рис. 7. Процесс детектирования лица каскадным детектором Multi-Task CNN

Все рассмотренные системы имеют два узких места. Первое: они последовательно применяют несколько независимых алгоритмов для получения конечного предсказания. Второе: если на изображении найдено несколько лиц, то алгоритм распознавания будет применен независимо к каждому детектированному лицу. Это приводит к большой потере производительности в силу избыточности вычислений. Время, которое каждый независимый компонент тратит на повторное вычисление похожих признаков, становится “бутылочным горлышком” для приложений. Особенно остро эта проблема встает в том случае, когда приложение должно работать в режиме реального времени.

Чтобы решить эту проблему, предлагается применить одношаговую архитектуру, используемую в задачах детектирования объектов. Такая архитектура строит отображение непосредственно от значений пикселей изображения к координатам ограничивающих прямоугольников и вероятностям классов. Такой подход может значительно сократить временные затраты.

**Постановка задачи**

Целью данной работы является создание системы автоматического детектирования и идентификации лиц на фотографии. К системе выдвигаются следующие требования: работа в режиме реального времени, корректная обработка фотографий с несколькими лицами в кадре, возможность легкой имплементации специализированных для задачи распознавания лиц техник. В качестве базового алгоритма используется сверточная нейронная сеть.

**Этапы работы**

* Сформировать базу данных из публично доступных наборов фотографий.
* Реализовать предобработку изображений.
* Разделить базу данных на обучающую, валидационную и тестовую выборку.
* Выбрать архитектуру и реализовать нейронную сеть.
* Обучить сеть и проверить ее работу на тестовом множестве.

**Теоретическая часть**

**Сверточные нейронные сети**

Сверточная нейронная сеть – особый тип нейронных сетей, который эффективным образом обрабатывает данные с пространственной структурой, в первую очередь изображения. Это достигается использованием операции свертки

Для сверточной нейронной сети, состоящей из L слоев, первые L−1 слоев могут быть описаны следующим образом:

где это обучаемые параметры сети. – это тензор весов двумерной свертки, – вектор смещений, – фиксированная нелинейная функция активации (обычно ReLU), применяемая поэлементно.

Последний слой производит непосредственную классификацию, т. е. переводит полученный тензор признаков в вероятности классов. Для этого в качестве функции активации используется функция softmax

**Пакетная нормализация и групповая нормализация**

Чтобы повысить производительность и ускорить обучение нейронных сетей, в [10] был представлен метод нормализации значений на скрытых слоях нейронной сети, названный пакетная нормализация (англ. batch-normalization). Он основывается на том наблюдении, что, с одной стороны, нормализация входных значений нейронной сети помогает улучшить обучение. С другой стороны, нейронная сеть является последовательностью слоев, в которой выходные значения каждого слоя являются входными для следующего. Поэтому каждый слой сети может рассматриваться как первый слой меньшей подсети, и нормализация промежуточных значений, получаемых на скрытых слоях нейронной сети, значительно улучшит обучение.

О конкретных теоретических обоснованиях работы пакетной нормализации до сих пор ведутся споры. Широко распространено мнение о том, что эффективность пакетной нормализации связана с уменьшением внутреннего ковариационного сдвига. Недавние исследования [11] склоняются к тому, что причина эффективности в сглаживании рельефа оптимизируемой функции.

Пакетная нормализация обычно проводится после применения операции свертки, но до применения функций активации. Она состоит из собственно нормализации и аффинного преобразования с обучаемыми параметрами:

Среднее *M* и дисперсия *D* вычисляются поканально по входному набору изображений.

В [12] было представлено улучшение данной техники под названием групповая нормализация. В групповой нормализации каналы входного тензора разбиваются на несколько групп, и среднее и дисперсия вычисляются по каждой группе отдельно.

**Архитектуры, основанные на регрессии-классификации**

Вместо того чтобы использовать многошаговую архитектуру, было предложено напрямую учить отображение из значений пикселей в значения координат ограничивающих прямоугольников и вероятности классов.

Одна из наиболее эффективных реализаций этой идеи представлена в сверточной сети YOLO [13]. Эта сеть делит изображение на сетку S × S, и каждая клетка данной сетки ответственна за предсказание объекта с центром внутри данной клетки. Каждая клетка сетки предсказывает *B* ограничивающих прямоугольников и соответствующее им значение уверенности. В то же время, каждая клетка способна детектировать лишь один объект. Вне зависимости от количества ограничивающих прямоугольников, каждая клетка сетки предсказывает единственный набор условных вероятностей классов .Вероятности классов для каждого ограничивающего прямоугольника вычисляется как произведение условной вероятности классов в данной клетке на значение уверенности данного ограничивающего прямоугольника

Из полученных ограничивающих прямоугольников отбрасываются те, у которых значение уверенности оказывается меньше заранее заданного порога. Затем производится процедура подавления немаксимумов, т. е. из нескольких предсказаний одного класса, ограничивающие прямоугольники которых пересекаются и значение IoU (intersection-over-union) для которых больше порога, остается только предсказание с максимальным значением уверенности.

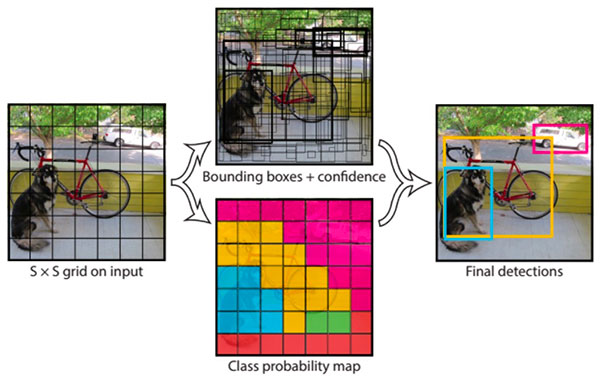


Рис. 8. Схема работы сети YOLO

Оригинальная архитектура обладала рядом недостатков, таких как ограничение на количество объектов в кадре и их близость друг к другу, плохое детектирование маленьких объектов и объектов, у которых сильно меняется соотношение сторон. Дальнейшие версии этой архитектуры – YOLOv2 [14] и YOLOv3 [15] – имеют несколько нововведений, которые призваны исправить эти недостатки.

Во-первых, вместо того, чтобы предсказывать произвольные ограничивающие прямоугольники, используются несколько заранее заданных априорных ограничивающих прямоугольников (англ. anchor box). Они представляют собой априорную информацию о размерах и форме детектируемых объектов. Размеры этих прямоугольников могут быть получены из интуитивных соображений или с помощью кластеризации (например, методом K средних) размеров ограничивающих прямоугольников из обучающей выборки.

Во-вторых, вместо непосредственного предсказания координат центра и размеров, предсказывается отступ центра соответствующего априорного ограничивающего прямоугольника от верхнего левого угла клетки и коэффициенты экспоненциального преобразования, переводящего размеры априорного ограничивающего прямоугольника в размеры предсказываемого ограничивающего прямоугольника.

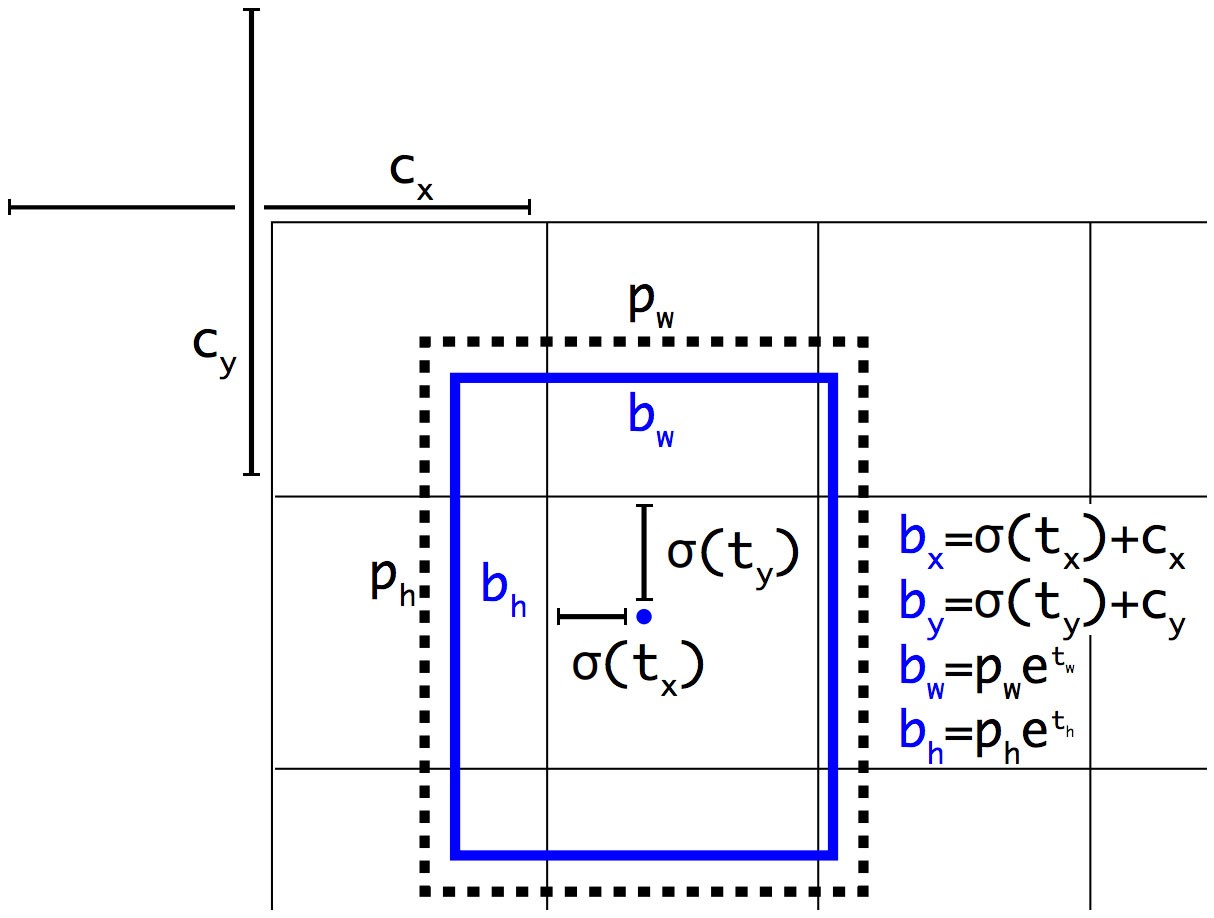


Рис. 9. Получение атрибутов ограничивающего прямоугольника из предсказанных сетью значений tx, ty, tw, th

Вероятности классов предсказываются напрямую для каждого априорного ограничивающего прямоугольника.

Для обучения такой сети используется особым образом построенная функция потерь. Она представляет собой сумму трех функций потерь: функций потерь локализации (для координат и размеров ограничивающего прямоугольника), функций потерь уверенности и функций потерь классификации.

Функция потерь локализации и функция потерь классификации вносят вклад в общую функцию потерь только в тех позициях сетки, где действительно находится объект. Более того, из нескольких предсказанных ограничивающих прямоугольников в каждой позиции сетки выбирается только один – тот, который имеет наибольший IoU с истинным ограничивающим прямоугольником.

**Практическая часть**

**Обучающая выборка**

Обучающая выборка составлена из наборов данных FEI Face Database [16], Caltech Faces 1999 [17], Georgia Tech Face Database [18]. Она представляет собой 3277 размеченных фотографии с одним из 285 субъектов в кадре. Наличие лишь одного человека в кадре является важным фактом. Большинство размеченных наборов данных для распознавания лиц содержит именно такие фотографии. А задача сбора и разметки специализированного набора данных с несколькими лицами в кадре представляется слишком сложной в рамках данной работы.

**Предобработка**

Для того чтобы получить информацию о положении лица на фотографии, а именно относительные координаты центра и относительные размеры ограничивающего прямоугольника, применяется детектор лиц из библиотеки dlib [19]. Затем все изображения уменьшаются до 320 пикселей по меньшей стороне.

**Аугментация данных**

Для искусственного увеличения обучающей выборки используется аугментация данных. Используются стандартные техники аугментации из [7] – вырезание случайного фрагмента размера 320×320, случайное отражение по горизонтали, случайное изменение цвета. Для того, чтобы сохранить соответствие между изображением лица и координатами ограничивающего прямоугольника, эти техники модифицируются следующим образом:

* вырезаемый случайный фрагмент должен полностью включать соответствующий ограничивающий прямоугольник (координаты центра прямоугольника модифицируются соответствующим образом);
* если к изображению применяется отражение по горизонтали, то оно применяется и к координатам центра ограничивающего прямоугольника.

**Архитектура сети**

В силу специфики задачи распознавания лиц было принято решение модифицировать архитектуру сети YOLO. В частности, сеть DarkNet, используемая в оригинальной архитектуре для извлечения признаков, заменена на более распространенную сеть ResNet [20]. Также процесс получения атрибутов ограничивающих прямоугольников отделен от процесса получения вероятностей классов для того, чтобы в дальнейшем была возможность непосредственно работать с признаками лиц, по которым вычисляются вероятности классов.

Архитектура сети представлена на рис. 10.

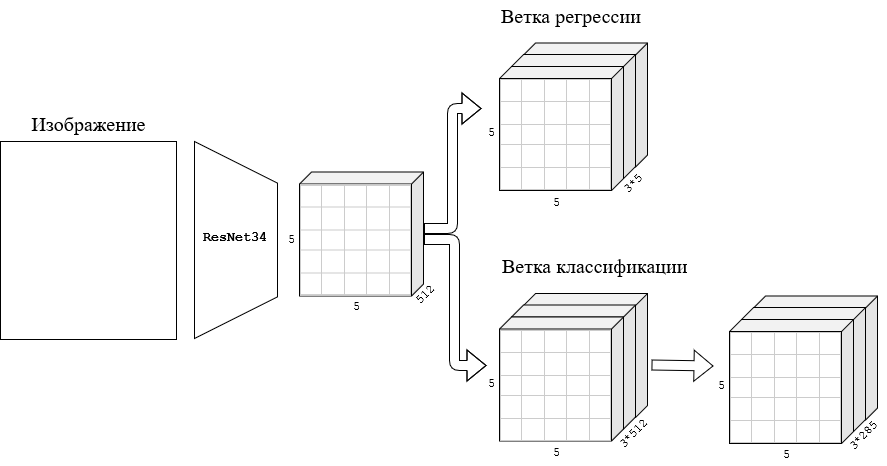


Рис. 10. Предлагаемая архитектура нейронной сети

Изображение поступает на вход сети ResNet34, отвечающей за извлечение признаков. Результирующий тензор представляет собой сетку 5×5, каждая клетка которой описывается вектором длины 512.

Нейронная сеть разделяется на две ветки. Ветка регрессии состоит из единственного сверточного слоя, результатом которого являются атрибуты ограничивающих прямоугольников, соответствующих 3 априорным ограничивающим прямоугольникам для каждой клетки сетки. Ветка классификации представляет собой следующую последовательность слоев: сверточный слой с последующей групповой нормализацией и функцией активации ReLU, сверточный слой с функцией активации softmax. Он получает на выходе тензор, содержащий вероятности классов для каждого из 3 априорных ограничивающих прямоугольников в каждой клетки сетки.

**Алгоритм обучения**

Для обучения данной сети используется функция потерь, которая основана на функции потерь сети YOLO. Для увеличения точности данная функция потерь модифицируется следующим образом. В качестве функций потерь локализации используется функция потерь Хьюбера (smooth L1 loss). В качестве функций потерь классификации используется кросс-энтропия, а в качестве функций потерь уверенности – бинарная кросс-энтропия.

**Реализация**

Нейронная сеть реализована [21] на фреймворке PyTorch. Обучение сети проводилось методом обратного распространения ошибки с использованием оптимизационного алгоритма Adam [22]. График изменения ошибки на обучающем и валидационном множестве представлен на рис. 11.

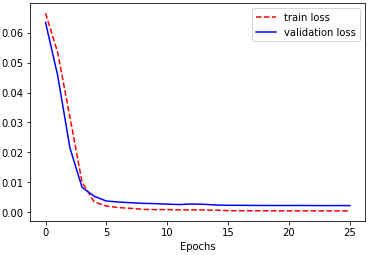


Рис. 11. График зависимости ошибки на обучающем и валидационном множестве от количества эпох обучения

**Тестирование сети и анализ результатов**

Для тестирования использовалась выборка, составленная из не вошедших в обучающую выборку изображений. Она состоит из 651 фотографии с одним из 285 субъектов в кадре.

В качестве базы для сравнения использовалась традиционная комбинация из детектора лиц библиотеки dlib и сети ResNet34 для классификации. Сравнивались следующие характеристики – доля правильных ответов, средний IoU, количество кадров в секунду.

Результаты сравнения представлены в таблице:

|  |  |  |
| --- | --- | --- |
|  | **dlib + ResNet34** | **Предлагаемая архитектура** |
| **доля правильных ответов** | 97,5% | 98,3% |
| **средний IoU** | 79,8% | 80,1% |
| **FPS** | 6,5 | 34,5 |

Видно, что сеть получает результаты на том же уровне при более чем пятикратном приросте скорости.

Несмотря на отсутствие в обучающей выборке фотографий с несколькими лицами в кадре, нейронная сеть обучилась хорошо обрабатывать данный случай.



Рис. 12. Результаты работы нейронной сети на фотографиях с несколькими лицами в кадре

**Заключение**

В работе предложено решение задачи распознавания лиц на фотографии. Приведена архитектура сверточной нейронной сети, решающей задачу детектирования и идентификации. Описаны методы предобработки, а также техники аугментации входных данных, которые позволяют эффективно обучить данную нейронную сеть. Показано, что в процессе обучения сети достигается хорошая точность, и при этом она обладает достаточной скоростью работы для удовлетворения требования работы в реальном времени. Архитектура сети предоставляет возможность непосредственно работать с признаковыми представлениями лиц, что является важным фактом для дальнейшего внедрения специализированных для задачи распознавания лиц техник.

**Список литературы**

1. Под присмотром: во сколько обойдется система распознавания лиц на улицах Москвы // Forbes URL: <https://www.forbes.ru/tehnologii/350843-pod-prismotrom-vo-skolko-oboydetsya-sistema-raspoznavaniya-lic-na-ulicah-moskvy> (дата обращения: 24.05.2019).
2. Zhang C., Zhang Z. (2010). A survey of recent advances in face detection URL: <https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/facedetsurvey.pdf> (дата обращения: 24.05.2019).
3. Beham M. P., Roomi S. M. M. A review of face recognition methods // International Journal of Pattern Recognition and Artificial Intelligence. 2013. Vol. 27(4).
4. Vaillant R., Monrocq C., LeCun Y. Original approach for the localisation of objects in images // IEE Proceedings: Vision, Image and Signal Processing. 1994. Vol. 141(4). P. 245–250.
5. Garcia C., Delakis M. Convolutional face finder: a neural architecture for fast and robust face detection // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004. Vol. 26(11). P. 1408–1423.
6. Osadchy M., LeCun Y., Miller M. Synergistic face detection and pose estimation with energy-based models // The Journal of Machine Learning Research. 2007. Vol. 8. P. 1197–1215
7. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classiﬁcation with deep convolutional neural networks // Communications of the ACM. 2017. Vol. 60(6). P. 84–90.
8. Taigman Y., Yang M., Ranzato M., Wolf L. DeepFace: closing the gap to human-level performance in face verification // IEEE Conference on Computer Vision and Pattern Recognition. 2014. P. 1701–1708.
9. Умные фотографии ВКонтакте URL: <https://vk.com/@td-highload-face-recognition> (дата обращения: 24.05.2019).
10. Ioffe S., Szegedy C. (2015). Batch normalization: accelerating deep network Training by reducing internal covariate shift URL: <https://arxiv.org/pdf/1502.03167.pdf> (дата обращения: 24.05.2019).
11. Santurkar Sh. Tsipras D., Ilyas A., Madry A. (2018). How does batch normalization help optimization? URL: <https://arxiv.org/pdf/1805.11604.pdf> (дата обращения: 24.05.2019).
12. Yuxin W., Kaiming H. Group normalization // Lecture Notes in Computer Science. 2018. P. 3–19.
13. Redmon J., et al. You only look once: unified, real-time object detection // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 779–788.
14. Redmon J., Farhadi A. YOLO9000: better, faster, stronger // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 6517–6525.
15. Redmon J., Farhadi A. (2018). YOLOv3: an incremental improvement URL: <https://arxiv.org/pdf/1804.02767.pdf> (дата обращения: 24.05.2019).
16. FEI Face Database URL: <https://fei.edu.br/~cet/facedatabase.html> (дата обращения: 24.05.2019).
17. Caltech data sets URL: <http://www.vision.caltech.edu/archive.html> (дата обращения: 24.05.2019).
18. Georgia Tech Face Database URL: <http://www.anefian.com/research/face_reco.htm> (дата обращения: 24.05.2019).
19. Frontal face detector // DLib URL: <http://dlib.net/imaging.html#get_frontal_face_detector> (дата обращения: 24.05.2019).
20. He K., Zhang X., Ren Sh., Sun J. Deep residual learning for image recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.
21. Код нейронной сети URL: <https://github.com/Andre6o6/face_detection_recognition> (дата обращения: 24.05.2019).
22. Diederik K., Ba J. Adam: a method for stochastic optimization // International Conference on Learning Representations (ICLR). 2015.