Филиал Московского Государственного Университета

имени М.В.Ломоносова в городе Ташкенте

Факультет прикладной математики и информатики

Кафедра прикладной математики и информатики

Нишанов Эмиль Рустамович

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

**на тему: «Распознавание эмоций по изображению лица человека» «Emotion recognition on human face images»**

**по направлению 01.03.02 «Прикладная математика и информатика»**

|  |
| --- |
| ВКР рассмотрена и рекомендована к защите  зав.кафедрой «ПМиИ», д.ф.-м.н., профессор\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Кудрявцев В.Б. |
|  |
| Научный руководитель  к.ф.-м.н., c.н.с. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Алексеев Д. В. |
|  |
| «\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ 2020г. |

Ташкент 2020 г.

**Содержание**

Аннотация . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 3

I. Введение . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1. Понятие «Эмоций» . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

2. Машинное обучение и эмоции . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5

3. Готовые решения . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6

II. Постановка задачи . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

1. Определение лица . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 11

2. Распознавание эмоций . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 15

III. Реализация . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

1. База изображений . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 16

2. Реализация на основе существующих методов. . . . . . . . . . . . . . . 17

3. Реализация уникальных методов . . . . . . . . . . . . . . . . . . . . . . . . . . 23

IV. Заключение. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 28

1. Перспективы . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

2. Благодарности . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 30

Список литературы . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 31

**Аннотация**

В данной выпускной квалификационной работе рассматривается построение системы автоматического распознавания эмоций. Предлагается обучение моделей сверточной неронной сети с учетом неравномерно распределенных по классам и относительно небольшого объема обучающих данных[7]. Целью работы является конструирование наилучшей модели в определении принадлежности в один из восьми классов (гнев, презрение, отвращение, страх, счастье, неопределенность, печаль, удивление).

**Abstract**

The construction of the automatic emotion recognition system is considered in this graduation qualification work. Convolutional Neural Network model training offered with unevenly distributed data between classes and relatively small volume dataset[7]. The aim of the work is to create the best model that defines membership in one of eight classes (anger, contempt, disgust, fear, happiness, neutral, sadness, surprise).

**I. Введение**

**1. Понятие «Эмоций»**

С рождения, каждый человек проявляет свое отношение к окружающему миру, образуя свое мировоззрение через разные формы познавательной деятельности. Общение с людьми, чтение литературы, выполнение работы, даже отдых могут вызывать разные чувства: радость, волнение, тревогу, злость. Все это разнообразные проявления отражательной психической деятельности человека, называемые эмоциями. Многие уверенно говорят об эмоциях, ставят гипотезы, пишут статьи и научные труды и, даже, целые книги по управлению и распознаванию переживаний и чувств. В последние два десятилетия значительно увеличилось количество исследований благодаря прикладным задачам. Однако, все еще нет согласия между экспертами в определении понятия эмоций. Например, автор книги «The Emotion Machine» Марвин Мински отмечает, что попытки психологов дать четкое определение словам больше помешали, чем помогли в создании теорий о том, как работает наш мозг. Сильное обобщение не дает ни единого шанса попыткам объединить результаты исследований и дать общие рекомендации. Слово “эмоции” трудно определить даже с толковым словарем. И чем больше словарей взять, тем больше возникает вопросов. Эмоции уверенно смешиваются с чувствами, состояниями, желаниями, аффектами. Попробуем дать определение термину «эмоции», воспользовавшись терминологией из психологического словаря[1].

**Эмоции** (от лат. emovere – возбуждать, волновать) – особый вид психических процессов или состояний человека, которые проявляются в переживании каких-либо значимых ситуаций (радость, страх, удовольствие), явлений и событий в течение жизни. Эмоции выступают в качестве главных регуляторов психической жизни и возникают в процессе практически любой активности человека. Эмоции возникли в процессе эволюции: с их помощью животные могли оценивать биологическую значимость явлений окружающего мира и внутреннего состояния организма.

Исследование в области психологии эмоций позволили далеко продвинуться в этой области, хотя еще в начале 20 века существовала достаточно фундаментальная теория о том, что не существует «универсальных» эмоций. Данная теория опиралась на зависимость представителя от типа культуры. На эту же тему в 1969 году была издана одна из самых важных литературных работ, проделанная психологами и учеными. Это исследование Пола Экмана и Уоллеса Фризена [2] [3], которые были явными оппонентами существующих канонов эмоционального происхождения. Они доказали, что существует 6 универсальных эмоций, которые не зависят от культурной и национальной принадлежности: удовольствие, страдание, страх, гнев, удивление и отвращение. Конечно же, к данным эмоциям всегда относят нейтральное состояние лица человека.

**2. Машинное обучение и эмоции**

Наука об эмоциях стала популярной не так давно, и в основном благодаря, выше названному, Полу Экману — американскому психологу, автору книги «Психология лжи» и консультанту популярного сериала «Обмани меня», который основан на материалах книги. Сериал стартовал в 2009 году, и в то же время значительно вырос публичный интерес к теме распознавания эмоций. Бум в стартап-среде случился в 2015-2016 годах, когда сразу два технологических гиганта — корпорации Microsoft и Google — доступными для обычных пользователей свои пилотные проекты для работы с наукой эмоций. Это стало толчком к созданию самых разных приложений и алгоритмов на базе технологии машинного обучения.

**Машинное обучение** (англ. machine learning, ML) [4][5] — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме. Благодаря машинному обучению программист не обязан писать инструкции, учитывающие все возможные проблемы и содержащие все решения. Вместо этого в компьютер (или отдельную программу) закладывают алгоритм самостоятельного нахождения решений путём комплексного использования статистических данных, из которых выводятся закономерности и на основе которых делаются прогнозы. Используя машинное обучение, компьютеры учатся распознавать на фотографиях и рисунках не только лица, но и пейзажи, предметы, текст и цифры.

**3. Готовые решения**

Автор[6] дает обзор готовых решений и методов, которые были успешно разработаны и внедрены в широкое пользование, а также существующие алгоритмы и методы, с помощью которых достигается решение основной задачи.

* **FaceReader** (Noldus Information Technology (Нидерланды))

Данное приложение может верно распознавать по выражению лица такие эмоции, как «счастье», «грусть», «злоба», «удивление», «страх», «отвращение» и «спокойствие» (Рисунок 1). Кроме того, FaceReader способен определить возраст, пол и расу людей. Приложение не нуждается в обучении и дополнительной настройке. В программе реализованы технологии компьютерного зрения. В частности, это метод Active Template, заключающийся в наложении деформируемого шаблона на изображение лица:

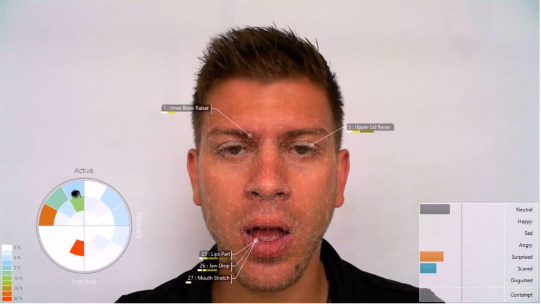


Рисунок 1-Интерфейс приложения FaceReader

Преимущества данной реализации:

1. Средний процент распознавания эмоций равен 89% и не зависит от наклона и поворота лица плоскости.
2. Программа работает с большим количеством форматов видеофайлов с кодеками MPEG1, MPEG2, DivX4, DivX5, DivX6, DV-AVI. При этом распознавание эмоций может происходить покадрово, либо полностью при просмотре всего видео. Приложение не ограничивается работой с загружаемым видео, FaceReader работает и со статичными изображениями и в реальном времени, считывая данные с камеры устройства пользователя.
3. Возможности визуализации: наличие гистограмм, диаграмм, отображение накладываемой сетки, процентов выражаемых эмоций.

Недостатки:

1. Процент распознавания эмоций и возраста сильно падает при работе с лицами детей возрастом меньше 5 лет.
2. Распознавание может происходить неточно, если человек носит очки.
3. Программа не определяет лица, повёрнутые в профиль.

* **EmoDetect** (Нейроботикс, Россия)

Программное обеспечение EmoDetect позволяет определить психоэмоциональное состояние человека по выборке изображений (видео или набор файлов). Разработанный классификатор эмоций позволяет определить 6 базовых эмоций: радость, удивление, грусть, злость, страх, отвращение. Определение эмоций происходит с помощью нейронных сетей: выполняется нахождение опорных точек на лице человека, и классификатор строит общую картину на основе их положения относительно друг друга. Компания Нейроботикс разрабатывает и поставляет системы для исследования физиологии человека и животных и, используя эти знания, разрабатывает новые технологии в области биомеханики, биоуправления и сенсорики в робототехнике.

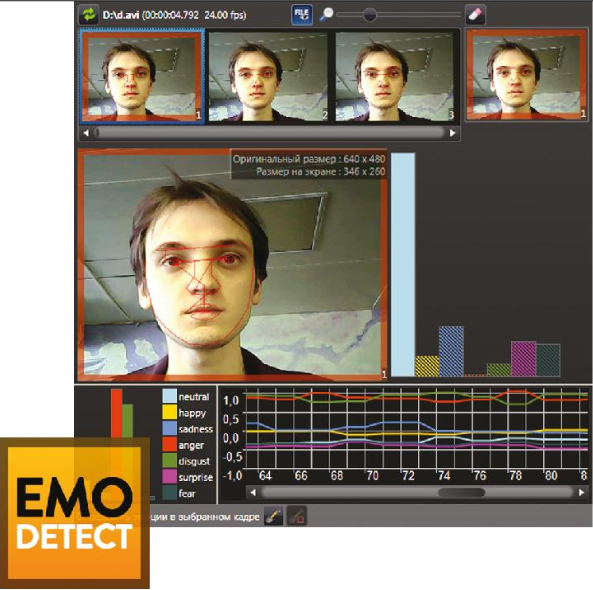


Рисунок 2 -Интерфейс приложения EmoDetect

Достоинства:

1. Классификация эмоций тремя независимыми классификаторами: нейронная сеть, система решающих правил, классификатор по взвешенной сумме признаков.
2. Построение графика динамического изменения интенсивности эмоций испытуемого во времени и формирование отчета о результатах обработки видео.
3. Возможность записи видео-файлов с веб-камеры, анализ всех поддерживаемых операционной системой форматов.

Недостатки:

1. Программа не определяет лица, повёрнутые в профиль.

* **FaceSecurity** ( Cognitec, Германия)

Данная реализация сканирует поток входящих данных и, при обнаружении в кадре лиц, сравнивает их с заведённой базой данных на предмет наличия совпадений. В случае нахождения совпадений система отправляет уведомления операторам в реальном времени. Данное приложение широко используется в системе безопасности европейских стран для выявления подозрительных лиц в общественных местах. Крупные компании же используют его для идентификации важных клиентов.Рядовые фирмы используют данную разработкудля распознавания лиц своих сотрудников (Рисунок 3) и выявления незарегистрированных посетителей, используя базы данных с фотографиями своих подчинённых. Также данная разработка помогает следователям идентифицировать лица в местах преступления по фотографиии видеонаблюдению путем сопоставления изображений лиц в хранилище агенств.

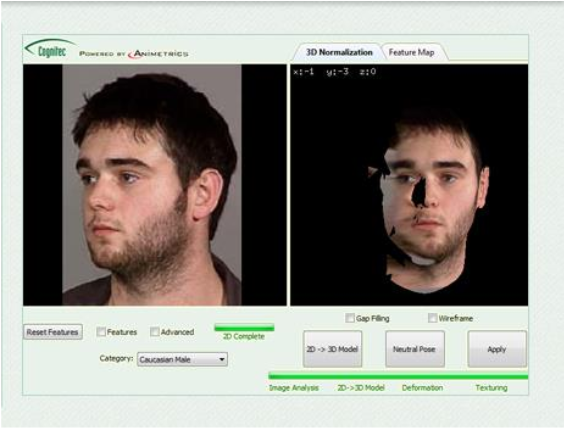


Рисунок 3 -Интерфейс приложения FaceSecurity

Достоинства:

1. Работа в реальном времени, возможность слежки за обнаруженным лицом на нескольких потоках данных одновременно.
2. В режиме реального времени лицо сравнивается с эталонами, занесёнными в базу данных.
3. Регистрации при неподвижном изображении или при живом потоке видео в ручном и автоматическом режиме.

Недостатки:

1. Невозможность произведения вычислений для лиц, имеющих отклонение от фронтального положения более чем 15 градусов.
2. Светочувствительность. Невозможность распознавания лица, при попадании на него теней и бликов.

**II. Постановка задачи**

Данная работа несет в себе исследовательский характер, так как задача была уже решена многими крупными компаниями, занимающиеся разработками систем по детектированию и распознаванию. Предлагается обучение ряд алгоритмов машинного обучения с учетом неравномерно распределенных по классам и относительно небольшого объема обучающих данных[7]. Целью работы является выбор наилучшей модели в определении принадлежности в один из восьми классов эмоций (гнев, презрение, отвращение, страх, счастье, нейтральность, печаль, удивление). Конечный алгоритм будет разделен на два основных последовательных этапа:

1. **Определение лица**

Определение (распознавание) лиц (Face Recognition - англ.) - это одни из наиболее перспективных методов биометрической бесконтактной идентификации человека по лицу. Первые системы распознавания лиц были реализованы как программы, устанавливаемые на компьютер. В наше время технология распознавания лиц наиболее часто используется в системах видеонаблюдения, контроля доступа, на разнообразных мобильных и облачных платформах.

В принципе, система распознавания лиц может быть описана как процесс сопоставления лиц, попавших в объектив камеры с базой данных ранее сохраненных и идентифицированных изображений лиц эталонов. Однако распознавание лиц состоит из множества взаимосвязанных подзадач:

* Во-первых, нужно посмотреть на картинку и найти на ней все лица.
* Во-вторых, нужно научиться распознавать каждое лицо, даже если оно странным образом повернуто, или если освещение плохое — ведь это все равно все тот же человек.
* В-третьих, нужно уметь определять уникальные черты лица, которые отличают одного человека от других, например, размер глаз, форма лица и так далее.
* Наконец, нужно сравнить выявленные уникальные особенности этого лица со всеми людьми, которых система уже знает, чтобы понять, кто изображен на фото.

Мозг человека делает всё мгновенно и незаметно для нас. Можно сказать, что люди настолько хороши в распознавании лиц, что видят лица даже в повседневных объектах. Компьютеры не способны к таким обобщениям высокого уровня (по крайней мере, пока …), поэтому нам придется учить их этому шаг за шагом.

Чтобы решить эту задачу, нужно построить своего рода конвейер, на котором мы выполняем каждый шаг распознавания по-отдельности, а затем передаем результат текущего шага следующему. Однако, для задачи распознавания эмоций достаточно первых два пункта подзадач.

Первое, что мы рассмотрим – это open-sourse библиотека OpenCV[8] (англ. Open Source Computer Vision Library, библиотека компьютерного зрения с открытым исходным кодом). Она представляет собой библиотеку программного обеспечения для компьютерного зрения с открытым исходным кодом и компьютерного обучения. Библиотека содержит более 2500 оптимизированных алгоритмов, которые имеют отличную точность в производительности и скорости.

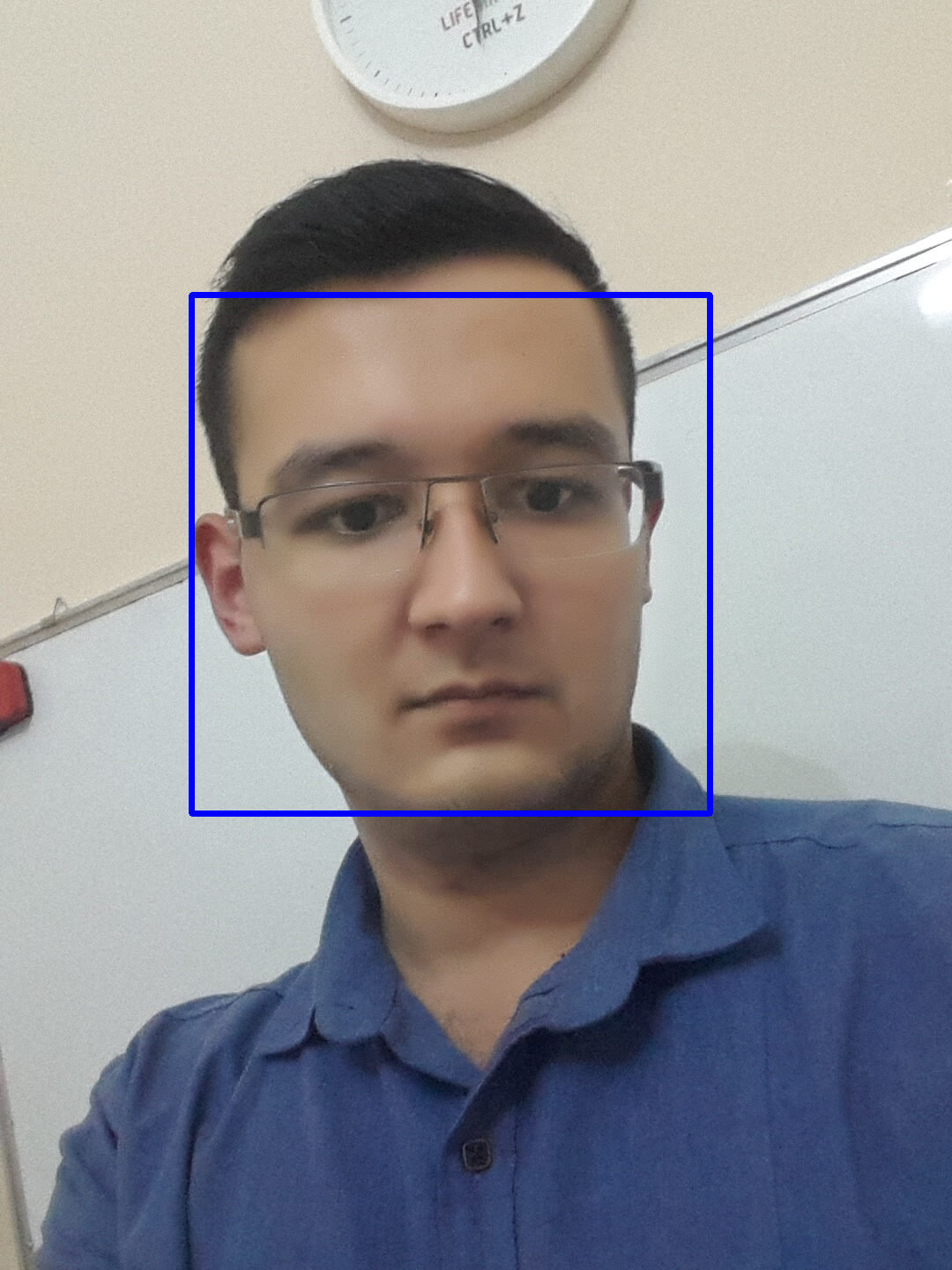


Рис. 5: Пример детектирования лица с помощью OpenCV

Данной библиотекой, использующую не безызвестный метод каскадов Хаара, активно пользуется разработчиками и молодыми учеными в сфере распознавания лиц в последние годы.

Каскады Хаара (Haar Cascade) [9] – это метод обнаружения объектов, используемый для определения местоположения объектов на изображениях. Алгоритм обучается на большом количестве положительных и отрицательных образцов: положительные образцы – это изображения, которые содержат интересующий объект, а отрицательные образцы – это изображения, которые содержат что угодно, кроме искомого объекта. После обучения классификатор может найти интересующий объект на новых изображениях.

Ключевой особенностью признаков Хаара является наибольшая, по сравнению с остальными признаками, скорость. При использовании интегрального представления изображения, признаки Хаара могут вычисляться за постоянное время (примерно 60 процессорных инструкций на признак из двух областей).

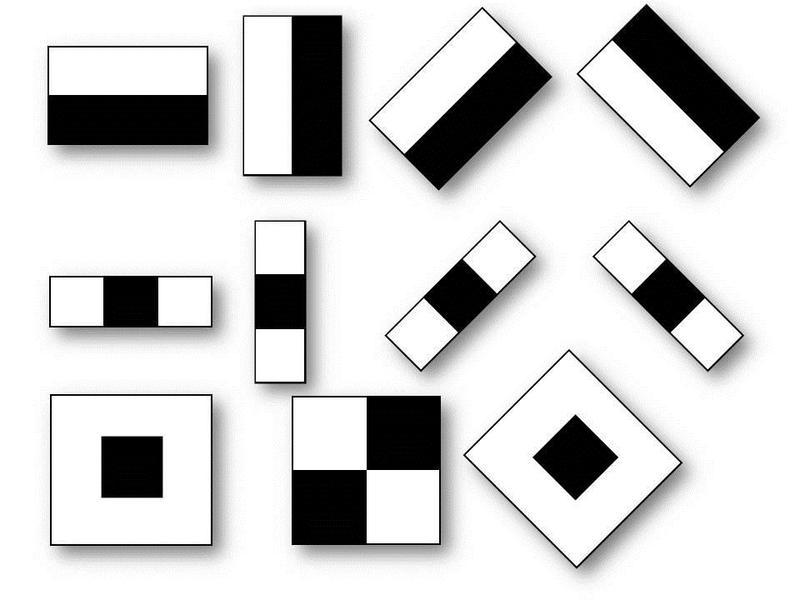


Рис. 6: Прямоугольные признаки Хаара

Для нашей задачи, распознавания эмоций, метод будет разделен на два этапа, как было сказано выше:

1. Нахождения лица на изображении;
2. Извлечение лица и приведение к стандартному, для задачи, размеру.

 b



а

Рис. 7: Пример распознавания лица и стандартизации изображения

1. **Распознавание эмоций**

Распознавание эмоций по выражению лица является сложным психическим процессом. Это умение является социальным навыком, приобретаемым естественным образом. Надо отметить, что для системы автоматического распознавания данная задача является куда более сложной, чем распознавание лиц.

Эта часть работы будет являться финальной и требует работы с полученными, в предыдущей части, данными. Как было упомянуто ранее, работа заключается в рассмотрении ряд известных алгоритмов и методов для сравнения и построении собственных методов для выбора наилучшей, с точки зрения приоритета нашей задачи, а именно, достижение максимальной точности на контрольной выборке из нашего множества данных[7].

Изображение

Распознаватель

Эмоция

Рис. 8: Схема нашей системы распознавания эмоций

**III. Реализация**

Как отмечалось в предыдущем разделе, реализация методов будет разделена на три части:

1. Рассмотрение методов и алгоритмов, которые были созданы в целях распознавания и детектирования, и использование их в задаче распознавания эмоций;
2. Создание собственных методов на базе полученных знаний и опыта из предыдущей работы, и текущего учебного года;
3. Сравнение и анализ рассмотренных методов и алгоритмов для выбора наилучшей.
4. **База изображений**

Первая проблема при проведениях исследований в области распознавания – это обучающая выборка. Обучающая выборка должна состоять из множества изображений, назовем ее M. Множество M состоит из объединения подмножеств, которые являются нашими классами эмоций (гнев, презрение, отвращение, страх, счастье, нейтральность, печаль, удивление). В идеале, наша выборка должна иметь достаточное количество изображений, примерно 10 000 на каждый класс, и с наименьшим количеством шума в изображении. Однако, составление такого набора требует значительных ресурсов, поэтому в качестве базы изображений, мной был выбран набор – facial\_expressions\_dataset[7]. Данная выборка содержит в себе 13 125 фотографий, разбитая на 8 классов и, в работе с ней, не нуждается в предварительной обработке, что позволяет сразу использовать в нашей задаче распознавания. Недостаток этой базы заключается в неравномерно распределенных классов. Так, к примеру, два класса – счастье и нейтральность покрывают практически 80% всей нашей выборки, что может повлиять на результат распознавания. Однако, именно такая выборка чаще всего отражает реальную ситуацию при анализе данных, что делает наше исследование более повседневной.

1. **Реализация на основе существующих методов (Transfer learning)**

В этой части работы рассмотрим методы, которые в последнее десятилетие стали самыми популярными в области компьютерного зрения, а, именно, сверточные нейронные сети[10][11][12]. На данный момент существует достаточно известных моделей сверточной архитектуры, которые успешно используется в коммерческой разработке компьютерного зрения, и мы также не останемся в стороне и применим метод Transfer learning.

Transfer Learning (TL) — одно из направлений исследований в машинном обучении, которое изучает возможность применения знаний, полученных при решении одной задачи, к другой.

Зачем же мы использовали Trasfer learning в нашей задаче распознавания эмоций?

1. Лишь немногие обучают сверточные сети с нуля (с помощью случайной инициализации), потому что зачастую нужные датасеты отсутствуют. Таким образом, использование предварительно обученных весовых коэффициентов или фиксированного экстрактора свойств помогает разрешить большинство проблем.
2. Сети с большим количеством слоев обучать дорого и долго. Самые сложные модели требуют нескольких недель для обучения с использованием сотен дорогостоящих графических процессоров.
3. Определение структуры сети, методов и параметров обучения — настоящее искусство, так как соответствующей теории, которая поможет разобраться с этими проблемами, просто не существует.

* **VGG16**

Чем же архитектура VGG16 [13] привлекла меня в ее использовании для задачи распознавания эмоций? Модель достигает точности 92.7% — топ-5, при тестировании на ImageNet в задаче распознавания объектов на изображении и одна из самых знаменитых моделей, отправленных на соревнование ILSVRC-2014. Она является улучшенной версией AlexNet, в которой заменены большие фильтры (размера 11 и 5 в первом и втором сверточном слое, соответственно) на несколько фильтров размера 3х3, следующих один за другим. Сеть VGG16 обучалась на протяжении нескольких недель при использовании видеокарт NVIDIA TITAN BLACK.

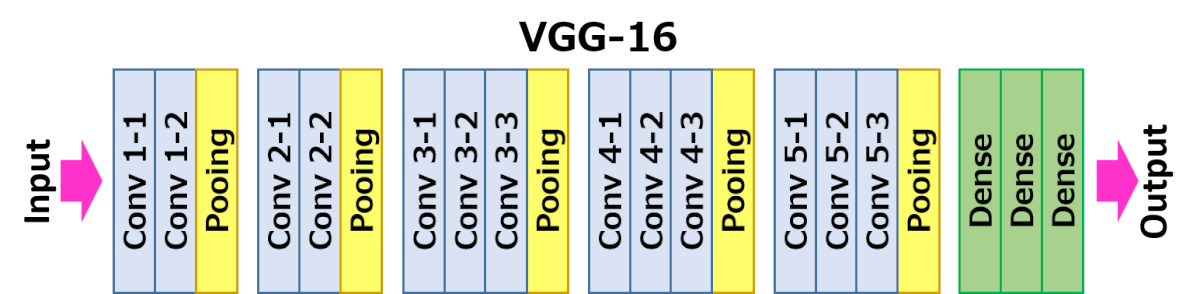


Рис. 9: Архитектура VGG16

В статье [14] подробно объяснено использование обученной части архитектуры, добавление собственной полносвязной части и обучение на собственном наборе изображении. При реализации я увидел, к сожалению, что сеть VGG имеет два серьезных недостатка:

1. Очень медленная скорость обучения.
2. Сама архитектура сети весит слишком много (появляются проблемы с диском и пропускной способностью)

Во избежание этих проблем, я воспользовался сервисом от компании Google для работы с машинным обучением, colab.research.google.com [15]. Благодаря облачным технологиям, скорость обучения и эффективность исследования превзошла мои ожидания.

Использование архитектуры VGG16 в десяти экспериментов показало, что сеть на имеющемся наборе достигает, максимум, 78% точности с минимальным значением потерь 0,915 на тестовой выборки из набора изображения. Для обучения оказалось достаточно 10 эпох с batch\_size, равным 500, после десятой эпохи не происходит никаких улучшений. На рисунке ниже показаны кривые значений точности и функции потерь при обучении.

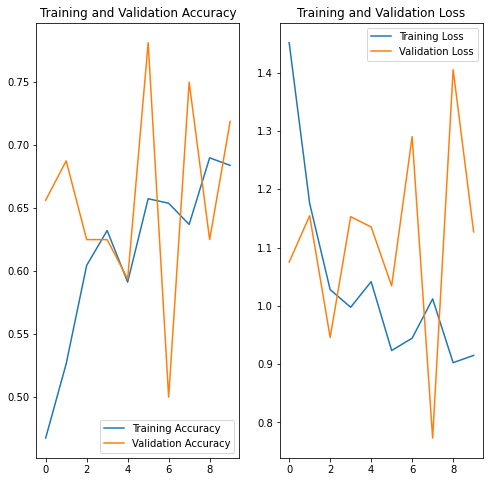


Рис. 10: ROC-кривая обучения архитектуры VGG16

Классификационная часть состояла из 5 слоев:

Dense(64, activation='relu') Dropout(0.3) Dense(32, activation='relu') Dropout(0.2) Dense(8, activation='softmax').

Оптимизатор: Adam(learning\_rate=0.001, beta\_1=0.9, beta\_2=0.999, amsgrad=False).

Пример работы на реальных фотографиях:



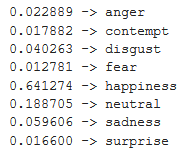


Рис. 11: Пример работы архитектуры VGG16

* **VGG19**

Следующая по очереди – это архитектура VGG19 [16]. Аналогична по построению предыдущей архитектуре, так как принадлежат семейству сетей архитектуры VGGNet, лишь с добавлением слоев свертки, делая модель более глубокой.

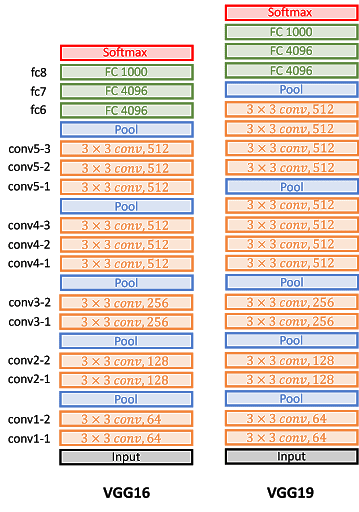


Рис. 12: Сравнение VGG16 и VGG19

Используя тот же классификатор, что и для предыдущей модели, можно ли сделать вывод, что чем глубже (сложнее) модель, чет лучше? Ответ на этот вопрос можно дать, совершив эксперимент.

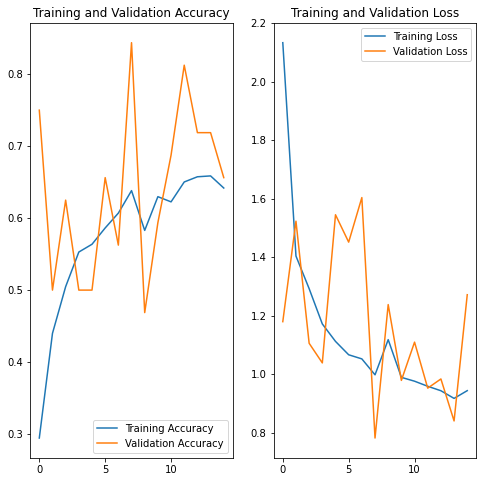
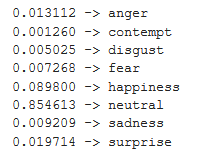


Рис. 13: ROC-кривая обучения архитектуры VGG19

Эксперимент показал наилучший результат 84% точности на контрольной выборке с значением функции потерь равным 0.9978. Однако, надо учитывать тот фактор, что выборка у нас распределена неравномерно. Из-за этого модель больше всего будет склоняться к выбору класса, где обучающий набор был больше.



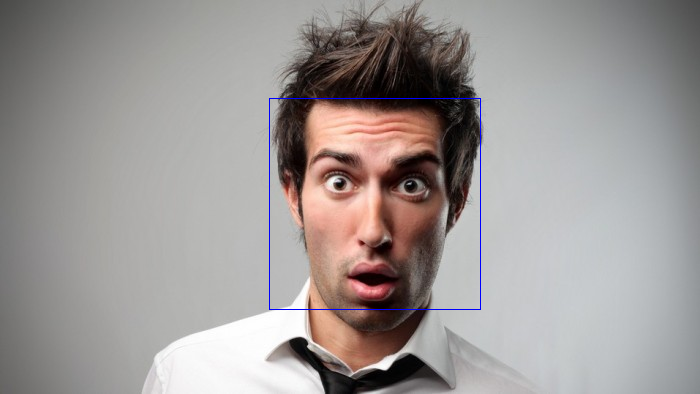


Рис. 14: Работа модели VGG19

На рис. 14 продемонстрировано, что нейрость ошибочно отнесла изображение лица в класс нейтрального состояния, хотя очевидно, что человек бы сказал, что лицо на изображении выражает удивление. Протестировав нейросети архитектуры семейства VGGNet, никаких выводов на счет использования метода Transfer Learning для задачи распознавания эмоций сделать невозможно, лишь можно подчеркнуть, что используемая выборка плохо подошла для данной архитектуры.

* **InceptionV3**

Последнее из архитектур в этой части – это **InceptionV3**, вышедшая в декабре 2015-го. Истоки этой модели исходят из архитектуры GoogLeNet. Количество параметров меньше, чем у VGGNet, но благодаря своему построению обходит своих конкурентов в соревновании ImageNet. На рис. 15 представлена архитектура сети.

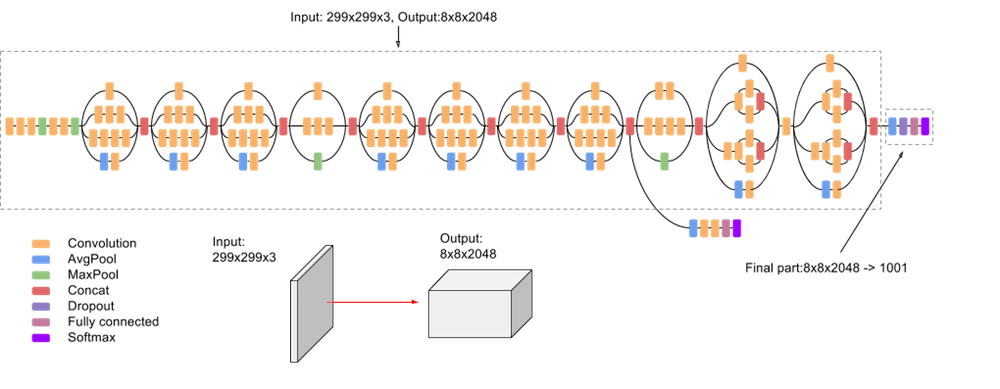


Рис. 15: Архитектура InceptionV3

Для задачи распознавания эмоций к InceptionV3 был обучен собственный классификационный слой. Модель обучалась 10 эпох, где наилучший результат на контрольном наборе был достигнут на 9-й эпохе с 62,5% точности и со значением функции потерь равным 1.3706.

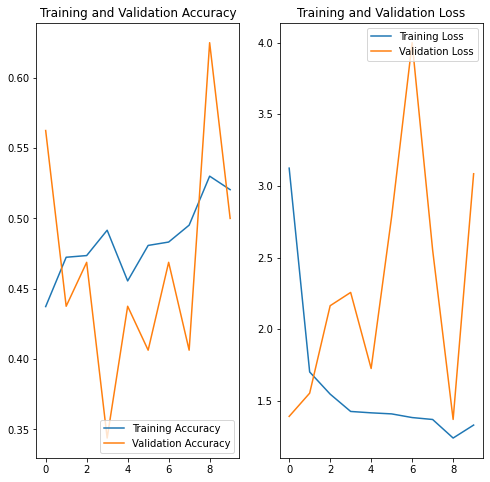


Рис. 16: ROC-кривая обучения архитектуры InceptionV3

1. **Реализация уникальных методов**

В предыдущей части, используя метод Transfer Learning, мы пытались ответить на вопрос об универсальном алгоритме, который способен работать и на других данных. В этой части попробуем найти и воссоздать методы, решающие задачу распознавание эмоций, используя наши данные [7].

* **AEI**

Автор статьи [17] рассматривает проблемы эмоционального искусственного интеллекта (Artificial ‘Emotional’ Intelligence) и описывает модель распознавания эмоции в видеопотоке. Модель состоит из 10 слоев сверток, 1 слоя подвыборки и 2 слоев нулевого дополнения. Дальше строится классификационная часть, состоящий из трех полносвязных слоев и на выходе с сигмоидальной функцией активации. Последнее связано с неравномерным набором данных, в статьях [18][19] даются рекомендации по работе с такими данными.

Модель обучалась 10 эпох, где наилучшим результатом по точности на контрольной выборке со значением 87,11% была достигнута на четвертой эпохе, хотя значение функции потерь на той же контрольной выборке понижалась вплоть до 6 эпохи, со значением 1.6485.

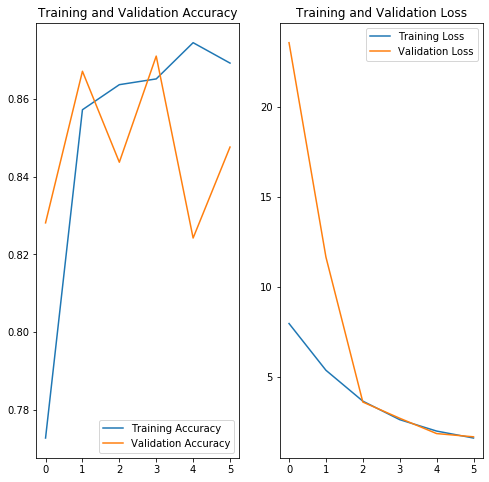


Рис. 17: ROC-кривая обучения

В отличии от известных архитектур, автор использовал менее емкую модель. Также впечатление произвело использование слоя свертки с ядром размера 2 на 2, как аналог слоя подвыборки maxpool/avaregepool. На рисунке ниже приведен пример работы сети на изображении не входящее в выборку.

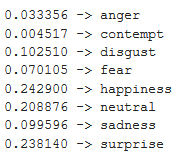




Рис. 18: Работа модели AEI

На изображениях, не входящих в выборку, модель при тестировании и анализе вела себя довольно интересно, показывая результаты, иногда, превосходящие мои ожидания. Однако модель чаще всего совершала ошибки на классах (отвращение, презрение и страх), где количество данных в выборке было самым наименьшим. Еще одним недостатком сети – это неустойчивость при элементарных преобразованиях объекта на изображении (зеркальное отражение, уменьшение/увеличение лица и пр.). Где под устойчивостью понимается минимальные изменения в предсказаниях при небольших изменениях (элементарных преобразованиях) объекта на изображении.

* **Собственная реализация**

Обучение нейросети — непредсказуемый и захватывающий процесс, который, однако, требует тщательной подготовки. В целом его можно разделить на три основных этапа:

* *Однократная настройка.* Сюда входят: выбор функции активации, предварительная обработка данных, инициализация весов, регуляризация, градиентная проверка.
* *Динамика обучения.* Отслеживание процесса обучения, оптимизация и обновление гиперпараметров.
* *Оценка.* Использование ансамблевых методов.

Создание же собственной (уникальной) архитектуры требует огромного опыта в этом направлении. Нет нигде описание или инструкции по созданию архитектур нейронных сетей, поэтому само создание можно ассоциировать с творческим подходом. В моей работе я также не остался в стороне и попытался воссоздать уникальный метод для решения задачи распознавания эмоций.

Модель, на момент написания работы, состояла из 8 сверточных слоев, 2 слоя подвыборки и 4 слоя пакетной нормализации. Далее следует классификационный слой, состоящий из трех полносвязных слоев и на выходе с сигмоидальной функцией активации. Последнее связано с неравномерно распределенной по классам выборкой. Основа при создании модели послужила предыдущая работа, посвященная распознаванию двух эмоций при небольшом наборе обучающихся данных.

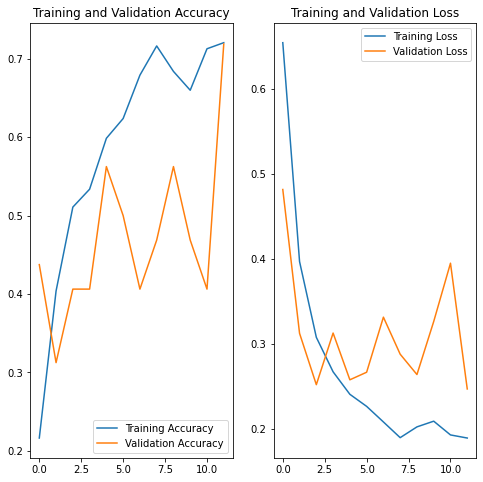


Рис. 19: ROC-кривая обучения

Было проведено десяток экспериментов, в основном связанных с тонкой настройкой модели, где под тонкой настройкой подразумевается подбор гиперпараметров сети. Финальная, на момент написания, модель обучалась 12 эпох, где наилучшим результатом по точности на контрольной выборке со значением 71,88% была достигнута на двенадцатой эпохе, со значением функции потерь 1.6485.

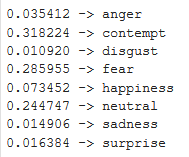




Рис. 20: Работа модели

Подбор гиперпараметров, по моему мнению, один из трудоемких процессов при обучении модели. Нужен огромный опыт при выборе значений гиперпараметра, ведь тонкая настройка может, как и улучшить, так и ухудшить способность сети адекватному обучению. При создании собственной модели я воспользовался регуляризации весов l1\_l2(0.01) и оптимизатором SGD(lr=0.01, decay=1e-6, momentum=0.9, nesterov=True).

Аналогичное тестирование на изображениях вне выборки прошло и над этой моделью. Результаты были довольно интересными, так как модель справлялась, по моему субъективному мнению, лучше, чем рассмотренные ранее модели, а именно на классах, где количество элементов в выборке было наименьшим. Хотя модель также неустойчива при элементарных преобразованиях объекта на изображении (зеркальное отражение, уменьшение/увеличение лица и пр.). На рис. 20 пример предсказания модели на реальном изображении.

После каждого эксперимента был проведен анализ модели на выявлении признаков на каждом сверточном слое. Как было сказано ранее, количество сверточных слоев в модели восемь.



Рис. 21: Пример анализа

**IV. Заключение**

В работе были рассмотрены методы распознавания эмоций по изображению лица человека. Как было сказано в введении, классификация эмоций другого человека — это субъективная оценка, которая вырабатывается на протяжении всей жизни. Поэтому к этой задаче было применено несколько подходов.

В начале был рассмотрен метод transfer learning, который заключался в использовании и адаптации известных архитектур сверточных нейронных сетей в нашей задаче. Результаты, по моему субъективному мнению, на реальных данных были весьма неудовлетворительными, что показывает нам неспособность одной архитектуры решать иные проблемы. Далее рассматривались иные подходы, а именно, уникальные методы, подстроенные под нашу задачу. В первой из них автор [17] в статье рассматривал проблемы «эмоционального» искусственного интеллекта, где предложил собственную архитектуру сверточной нейронной сети. Результаты показались мне весьма удовлетворительными, по сравнению выше названными моделями из метода transfer learning. Последняя модель, рассмотренная в нашей работе – это собственная архитектура сверточной нейронной сети, построенная на базе предыдущей моей работы. Аналогично ранее озвученной, сеть показала неплохие результаты, по сравнению с моделями из метода transfer learning, но уступает по точности модели, предложенной автором [17]. Одним из преимуществ в ходе экспериментов можно выделить скорость обучения и прогнозирования собственной архитектуры над остальными.

В работе были не упомянуты другие методы, полностью провалившиеся в нашей задаче. Они не включали в себя архитектуру нейронных сетей. Одним из методов для распознавания эмоций использовался алгоритм, в основе которого лежит использование фундаментальных статистических характеристик: средних (мат. ожидание) и ковариационной матрицы; использование метода главных компонент[20]. В качестве классификации использовался алгоритм Байесовского классификатора.

При выполнении практической и теоретической части работы были выполнены данные поставленные задачи:

* Изучение такого инструмента машинного обучения как нейронные сети, которые снова и снова доказывают свою эффективность в задачах анализа данных.
* Рассмотрение и применение готовых решений в задаче распознавания эмоций.
* Моделирование собственной архитектуры, которая, по моему субъективному мнению, не уступает в точности прогнозировании и имеет преимущество в скорости обучения и прогнозирования.

**1. Перспективы**

Завершая данную работу, было отмечено для себя множества перспектив развития данной области. В первую очередь – это усовершенствование алгоритмов, для более тонкого анализа в распознавании лицевых экспрессий. Далее можно заметить, что в распознавание входит не только зрительные, но и слуховые образы. Поэтому при конструировании системы распознавания эмоций нужно учитывать и голос человека, которого мы хотим распознать. Естественно не забывая, что эмоции – очень субъективная окраска человеческого состояния, которую понять очень часто не может сам человек, не используя комплекс иных деталей. Следовательно, в перспективе распознавание эмоционального состояния человека – создание комплекса алгоритмов, которые учитывали бы все аспекты.

Необходимость создания и интегрирования систем распознавания уже становится понятной многим крупных компаниям, так как влечет за собой безопасность, интерес государства и огромный финансовый успех на рынке систем безопасности. Также надо отметить, что исследования и развитие данной области ведет науку и технологии на создание более устойчивых интеллектуальных систем для различной формы деятельности. Ведь умение понять состояние человека – одна из уникальнейших способностей, пока в полной мере недоступна компьютерной программе.

**2. Благодарности**

Я выражаю глубокую признательность за помощью в подготовке выпускной квалификационной работы моему научному руководителю – Д. В. Алексееву. Его педагогический и научный подход, отличное знание предмета вдохновили меня на выполнение данной работы. Все время от выбора темы до завершения работы он помогал мне рекомендациями, советами, оказывал моральную поддержку.

Позвольте выразить искреннюю благодарность моим рецензентам за проделанную работу, рекомендации и замечания, а так же членам аттестационной комиссии, руководству университета, моим преподавателям.

Сердечная благодарность моим родным и близким за поддержку и помощь.

**Список литературы**

[1] http://www.psychologies.ru/glossary/27/emotsii/

[2] Ekman P., Darwin and Facial expressions, Academic Press, 1973

[3] Ekman P., Facial expression and emotion, American Psychologist, 48:384-392, 1993

[4] http://www.machinelearning.ru/wiki/index.php?title=Машинное\_обучение

[5] Pedro Domingos, The Master Algorithm, 2016

[6] Лескин Владимир Николаевич, Распознавание эмоций в видеопотоке, 2016

[7] https://www.kaggle.com/c/emotion-detection-from-facial-expressions/overview

[8] https://opencv-python-tutroals.readthedocs.io/en/latest/py\_tutorials/py\_tutorials.html

[9] http://wiki.amplab.ru/cgi-bin/awki.cgi/КаскадХаара

[10] https://habr.com/ru/post/348000/

[11] https://habr.com/ru/post/309508/

[12] https://habr.com/ru/post/436838/

[13] https://iq.opengenus.org/vgg16/

[14] https://machinelearningmastery.com/use-pre-trained-vgg-model-classify-objects-photographs/

[15] https://github.com/deepmipt/dlschl/wiki/Инструкция-по-работе-с-Google-Colab

[16] https://iq.opengenus.org/vgg19-architecture/

[17] https://towardsdatascience.com/aei-artificial-emotional-intelligence-ea3667d8ece

[18] https://towardsdatascience.com/handling-imbalanced-datasets-in-machine-learning-7a0e84220f28

[19] https://machinelearningmastery.com/cost-sensitive-neural-network-for-imbalanced-classification/

[20] https://habr.com/ru/post/68870/

[21] К. В. Воронцов, Математические методы обучения по прецедентам (теория обучения машин), 2015

[22] Орельен Жерон, Прикладное машинное обучение с помощью Scikit-Learn и Tensorflow, 2018