|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

Факультет «Информатика и системы управления»

Кафедра ИУ5 «Системы обработки информации и управления»

**РЕФЕРАТ**

**по дисциплине «Технология мультимедиа» на тему:**

**«*Методы и алгоритмы обнаружения улыбок на лицах на видео и фотографиях*»**

Выполнил студент: \_\_***Белоусов Евгений Александрович***\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*фамилия, имя, отчество*

Группа: \_\_\_\_***ИУ5-61Б***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Проверил доцент, к.т.н**\_\_\_\_\_\_\_\_\_\_\_\_Г.И. Афонасьев**

*подпись, дата*

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Дата \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*2020 г.*

# Оглавление

[Оглавление 1](#_Toc39974258)

[Введение 2](#_Toc39974259)

[Основная часть 3](#_Toc39974260)

[Задача обнаружения улыбок на лицах на видео и фотографиях в терминах машинного обучения 3](#_Toc39974261)

[Возможные решения задачи без использования нейронных сетей 5](#_Toc39974262)

[Нейросетевой подход для решения задачи обнаружения улыбок на изображении 7](#_Toc39974263)

[Существующие общепринятые решения 9](#_Toc39974264)

[Вывод 12](#_Toc39974265)

[Литература 13](#_Toc39974266)

# 

# Введение

В современном мире существует много задач, которые раньше могли решаться только человеком. Однако развитие технического прогресса не стоит на месте и появляются алгоритмы, позволяющие решать их компьютерам или даже телефонам. Соответственно перед вычислительными системами ставятся новые более сложные задачи, требующие реализовывать такие алгоритмы.

Одним из примеров таких алгоритмов являются алгоритмы машинного обучения [2]. Появление этих алгоритмов вместе с увеличением вычислительных мощностей компьютеров привело к тому, что задачи, которые раньше мог решать только человек стали успешно решаться машинами. Например, предсказание стоимости товара в зависимости от его характеристик или оценивание шансов на победу определенной команды в матче – задачи, в которых активно используется машинное обучение.

Алгоритмы машинного обучения также получили широкое распространение в задачах анализа и обработки изображений. Как правило, для этих целей используются нейронные сети [3] различных архитектур.

В этой работе мы рассмотрим возможные методы и алгоритмы для решения задачи обнаружения улыбок на лицах на видео и фотографиях. Для этого мы:

1. рассмотрим задачу обнаружения улыбок в терминах машинного обучения;
2. представим возможные решения задачи без использования нейронных сетей;
3. рассмотрим нейросетевой подход для решения данной задачи;
4. посмотрим существующие общепринятые решения.

# Основная часть

## Задача обнаружения улыбок на лицах на видео и фотографиях в терминах машинного обучения

В качестве обучающих данных мы будем использовать фотографии с изображением улыбающихся и неулыбающихся людей. Каждый элемент обучающей выборки состоит из набора пикселей и метки, определяющий наличие улыбки на данной фотографии.

Перед началом работы с картинкой необходимо, чтобы все картинки были одного размера. Это можно достичь, например, с помощью применения функции pyrUp или pyrDown из библиотеки OpenCV [2].

Также, в зависимости от алгоритма, может потребоваться преобразование цветной картинки к черно-белой или даже бинарной. Это легко осуществляется с помощью изменения параметров функции imread.

После обучения на обучающем наборе, будем проводить тестирование на тестовом наборе данных, который представляет собой то же, что и обучающий, но с новыми, не встречающимися до картинками и неизвестными для алгоритма метками наличия улыбок.

Так как перед нами стоит задача определения наличия или отсутствия улыбки на картинке, по сути, мы должны разделить все элементы нашего набора данных на два класса: «улыбка присутствует» и «улыбка отсутствует». В связи с этим в данной работе будем рассматривать различные алгоритмы классификации [4].

Но как же быть с видео? Видеофайл состоит из набора изображений и аудио ряда. Мы можем разбить видео на отдельные картинки и применить к ним такие же действия, как и к отдельным фотографиям [5]. Дальнейшие действия будут зависеть от конкретной поставленной задачи. Например, если мы составляем программу для отлова улыбок, чтобы после сфотографировать улыбающегося человека и выписать ему штраф, то покадровая обработка нас вполне устроит.

Если же нам надо установить факт наличия улыбки, например, в короткометражном фильме, то мы можем применить покадровую обработку к каждому кадру видеозаписи, а после принять решение о наличие улыбки в видео по проценту кадров фильма, содержащих улыбку. Такой метод будет более устойчив к шуму, его точность будет выше, однако полнота меньше.

## Возможные решения задачи без использования нейронных сетей

Сегодня, пожалуй, что повсеместно для анализа изображений используются нейронные сети. Однако, давайте начнем наш обзор с более простых алгоритмов машинного обучения. Случаются ситуации, когда более простой и дешевый алгоритм (с точки зрения вычислительных и временных затрат) показывает более высокую точность в сравнении с более сложными и дорогими алгоритмами. Поэтому не будем сбрасывать простые алгоритмы со счетов и рассмотрим их в этой главе.

Определение наличия улыбки на фотографии – сложная задача, которая зависит от большого количества взаимного признаков (пикселей) картинки. В связи с этим, мы считаем, что нет смысла рассматривать линейные алгоритмы для решения задачи анализа изображений. Так как наличие или отсутствие улыбки на фотографии в первую очередь связано с расположением определенных пикселей относительно друг друга, а не расположением на картинке относительно картинки (центра изображения, краев и т.п.…).

В связи с такими же соображениями не будут работать метрические алгоритмы. Скорее всего они смогут выявить похожих людей, но не факт наличия улыбки на картинке. Это связано с тем, что для метрических алгоритмов не имеет значения, какие именно пиксели совпадают или различаются с образцом из обучающей выборки. Имеет значение только сам факт наличия или отсутствия совпадения. Так как на большинстве фотографий улыбка занимает сравнительно небольшой ее участок, скорее всего данный метод будет отдавать большие предпочтения совпадению цвета кожи или форме лица.

Еще одним популярным алгоритмом в машинном обучении являются случайные деревья и ансамбли алгоритмов [6], состоящие из них. Эти алгоритмы являются более сложными, нежели линейные модели, и поэтому их несомненно стоит протестировать при построении модели для конкретной задачи, однако они и не столь сложные, как нейронные сети, поэтому скорее всего не смогут точно выцепить уж очень сложную зависимость, коей является наличие улыбки на фотографии.

В следующем разделе рассмотрим популярные подходы к задаче обнаружения улыбки на фотографии с применением нейросетевых технологий.

## Нейросетевой подход для решения задачи обнаружения улыбок на изображении

Нейронная сеть – очень популярный и сложный метод машинного обучения. Пожалуй что, когда мы слышим как кто-то говорит «нейронная сеть», первая модель, которая приходит на ум – персептрон [7]. Персептрон – одна из самых простых моделей, которая имеет множество реализаций в различных библиотеках. При этом она может выискивать достаточно сложные зависимости в данных и должна показать неплохое качество на тесте.

Более сложным вариантом является сверточная нейронная сеть [8]. В ее основе находится операция свертки – поэлементное умножение каждого фрагмента изображения на матрицу (ядро) свертки, последующее суммирование и запись результата в соответствующую позицию. Такая операция повторяется много раз чередуясь с другими, похожими на нее, что приводит к тому, что сверточная нейронная сеть обычно состоит из большого числа слоев, поэтому их также часто называют глубокими (глубинными) нейронными сетями. Сверточная нейронная изначально создавалась для работы с изображениями, поэтому она без проблем может работать с цветным изображением без каких-либо трюков со стороны разработчика. На данный момент разработано много архитектур сверточных нейронных сетей эффективно справляющихся с анализом изображений, среди них – AlexNet, VGG, GoogleNet и ResNet. Чтобы получить хороший результат нужно выбрать одну из предложенных архитектур, изменить выходные слои (чтобы классифицировать объекты на 2 класса) и обучить ее обучающей выборке.

В результате получится модель, эффективно справляющаяся с поставленной задачей, которая скорее всего будет удовлетворять нашим требованиям. Однако, может случиться ситуация, когда на фотографии окажутся и уши, и улыбка, и нос, и глаза (все составляющие человеческого лица), но на картинке они будут находиться хаотично относительно друг друга. Например, сверху будет один глаз, ниже него ухо, левее от уха – улыбка, а под улыбкой – нос. В этом случае сверточная нейронная сеть не заметит подвоха и распознает наличие улыбки.

Решить данную проблему может использование капсульной нейронной сети. Капсульная нейронная сеть была задумана в 1979 году Джеффри Хинтоном и опубликована осенью 2017 [9].

## Существующие общепринятые решения

В предыдущих разделах мы описывали алгоритмы, которые могут быть использованы для анализа изображений на наличие улыбки. В этом разделе, рассмотрим существующие готовые решения по этой теме.

Анализ изображений – довольно специфическая задача. Для ее решения существуют специальные библиотеки, содержащие алгоритмы для работы с картинками. Эти библиотеки совмещают в себе алгоритмы машинного обучения, численные алгоритмы и алгоритмы для обработки изображений.

Одной из самых популярных таких библиотек является OpenCV. Библиотека OpenCV – библиотека с открытым исходным кодом. Библиотека может работать с большим количеством языков, среди которых Python и C++. Все вычисления над тензорами и матрицами в OpenCV реализованы с помощью библиотеки NumPy, которая является очень высоко оптимизированной, следовательно, вычисления, проведенные с ее помощью, будут проводиться быстрее. Это может оказаться важным при работе с видео или при обработке большого количества фотографий.

Так как мы будем искать улыбки на лицах людей (предположим, что улыбающиеся животные нас не интересуют) стоит обратить внимание на каскад Хаара [10]. Этот алгоритм был предложен Полом Виолой и Майклом Джонсом в 2001 году. Каждый алгоритм каскада основан на ансамбле простых алгоритмов, обученных на большом количестве отрицательных и положительных образов. При этом каждый алгоритм сокращает интересующую нас область, которую нам стоит проверять в дальнейшем.

Таким образом, используя каскад Хаара из OpenCV мы можем не проверять на наличие улыбки всю фотографию, а выделить область изображения, на которой находится лицо, затем область, на которой находится рот, и только затем, после очередного масштабирования, применять алгоритмы машинного обучения для того чтобы уже исключительно по губам определить, улыбается человек или нет.

Такой подход приведет к увеличению качества классификации и позволит сильно сократить обучающую выборку.

Еще одним перспективным подходом, основанным на использовании специальных библиотек, является использование библиотеки dlib. Эта библиотека во многом похожа на библиотеку OpenCV, но является намного менее популярной. Однако в ней представлены некоторые технологии, реализаций которых нет в OpenCV.

Одна из таких технологий – поиск особых точек. Эта технология позволяет выявить контуры лица. В дальнейшем из них можно попытаться вычленить контур рта и использовать в качестве входных данных для нейронной сети уже не все изображение рта, а только его контур. Тем самым мы сократим количество входных данных, сохранив при этом интересующую нас информацию.

Также, особые точки используются для трансформации лиц. Большинство открытых наборов данных для распознаванию эмоций содержат только фронтальные фотографии лиц. Если нам не гарантируется, что все изображения, которые будет анализировать наша модель, будут фронтальными, то мы попадем в неприятную ситуацию. Первый вариант ее решения – составить свой датасет. Для качественного датасета требуется большое количество данных, нам придется сделать много тысяч фотографий разных людей.

Другой вариант решения данной проблемы – привести изображение лица к фронтальному состоянию. Данная методика позволит сделать наш классификатор еще точнее и сократить количество ошибок.

Также, при использовании описанных выше решений, нам потребуется меньшая обучающая выборка для обучения нейронных сетей, что положительно скажется на стоимости разработки системы. Однако, дополнительные преобразования могут привести к тому, что время работы алгоритма на одном изображении будет дольше, и данный алгоритм перестанет устраивать нас при обработке видео.

# Вывод

В данной работе мы рассмотрели основные подходы к обнаружению улыбки на фотографиях.

Были рассмотрены преобразования, которые стоит провести над изображением до начала поиска улыбки на нем.

Был описан подход к поиску улыбки на видео.

Также были рассмотрены основные методы и алгоритмы, используемые непосредственно для обнаружения улыбки на картинке. Скорее всего, лучше всего будут работать сложные глубокие или даже капсульные нейронные сети, однако во время решения не стоит забывать и о простых алгоритмах машинного обучения.

Таким образом, для решения задачи обнаружения улыбки лучше всего использовать специальную библиотеку (такую как OpenCV) и одну из реализаций нейронных сетей глубокого обучения.

Также не стоит забывать, что конкретный алгоритм будет существенно различаться в зависимости от конкретной задачи.

# Литература

1. [https://docs.opencv.org](https://docs.opencv.org/)Жерон, Орельен. Прикладное машинное обучение с спомощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. : Пер. с англ. - СпБ. : ООО «Альфа-книга», 2018. -688с : ил. - Парал. тит. англ. ISBN 978-5-9500296-2-2 (рус)
2. OpenCV 3.0.0-dev documentation. Ссылка - https://docs.opencv.org/3.0-beta/index.html (ссылка от 7.05.2020)
3. *Беркинблит М. Б.*  Нейронные сети. — : МИРОС и ВЗМШ РАО, 1993. — 96 с.
4. *Вапник В. Н., Червоненкис А. Я.* Теория распознавания образов. — М.: Наука, 1974.
5. *Л. Шапиро, Дж. Стокман.* Компьютерное зрение = Computer Vision. — : Бином. Лаборатория знаний, 2006. — 752 с. — ISBN 5-94774-384-1.
6. *Hastie, T., Tibshirani R., Friedman J.* Chapter 15. Random Forests // The Elements of Statistical Learning: Data Mining, Inference, and Prediction. — 2nd ed. — Springer-Verlag, 2009. — 746 p. — ISBN 978-0-387-84857-0
7. *Ежов А. А., Шумский С. А.* Нейрокомпьютинг и его применения в экономике и бизнесе. ИНТУИТ (2006).
8. *Гудфеллоу Я., Бенджио И., Курвилль А.* Глубокое обучение = Deep Learning. — : ДМК Пресс, 2017. — 652 с. — ISBN 978-5-97060-554-7.
9. Geoffrey E. Hinton, Sara Sabour, Nicholas Frosst Dynamic Routing Between Capsules 2017
10. P. Viola and M. Jones. Robust real-time face detection. IJCV 57(2), 2004