**РАЗРАБОТКА СИСТЕМЫ РАСПОЗНАВАНИЯ ЭМОЦИЙ**

***Коровин Е.А.***

***Национальный исследовательский университет «Высшая школа экономики»,***

***департамент электронной инженерии***

***МИЭМ НИУ ВШЭ***

**Аннотация**

В исследовательской работе представлен алгоритм распознавания эмоций. Был проведен обзор существующих методов и средств для решения подобных задач. Для достижения поставленной цели использовались инструменты библиотеки OpenCV в языке программирования Python 3. Результаты экспериментов подтверждают конкурентоспособность разработанной модели в сравнении с аналогичными решениями, а точность распознавания эмоций превышает 68%.

**Введение**

Прогресс в области информационных технологий, внедрение которых в повседневную жизнь обычных граждан продолжает наращиваться, оказывает значительное воздействие не только на внешние аспекты их жизнедеятельности, подразумеваемые термином "цифровизация", но также на их поведенческие характеристики. Этот вопрос стимулирует интерес к тому, как с использованием текущего технологического прогресса, обусловленного накоплением обширных данных и увеличением доступности индивидуальных средств взаимодействия с цифровой средой, возможно обрести новые методы понимания, взаимодействия и воздействия как на группы людей, так и на индивида.

В рамках таких исследований анализ данных, включая распознавание эмоций по изображению лица человека, приобретает все большую актуальность. Это обусловлено увеличивающимся интересом к пониманию человеческих эмоций как в сфере маркетинга и рекламы, так и в психологии, медицине и других дисциплинах.

Исследования в области распознавания эмоций обладают значительным потенциалом для практического применения в различных областях. Например, в медицине они могут быть полезны для диагностики психических расстройств, анализа эмоционального состояния пациентов и мониторинга их прогресса. В области маркетинга и рекламы они способны улучшить понимание реакций потребителей на продукты и услуги, помогая компаниям адаптировать свои стратегии к их потребностям и, таким образом, оптимизировать коммуникацию между ними.

**Описательная постановка цели и задачи**

Целью данной работы является разработка системы распознавания эмоций по фотографии. В ходе анализа поставленной задачи, была выявлена необходимость создания полноценного многоэтапного алгоритма, который включал бы в себя не только определение эмоции по фото лица человека, но и этапы поиска на изображении лиц, выделение каждого лица в отдельности и предоставление результатов распознавания эмоции по каждому из найденных лиц. В качестве требований к разрабатываемому решению выдвигается: показатель точности распознавания, который должен являться сопоставимым или превышающим существующие публичные решения, а время распознавания, что не должно превышать одной секунды.

Процесс распознавания эмоций представляет собой последовательный набор этапов. На первом этапе осуществляется локализация лица на изображении, где определяется точное положение лица среди других объектов на фотографии или видеозаписи.

Далее следует этап нормализации изображения лица, где происходит стандартизация изображения с целью устранения возможных искажений, таких как повороты или изменения освещения, что позволяет достичь более точного и унифицированного представления лица для последующего анализа.

Последний этап – классификация, где производится определение к какому из определенных классов эмоций относится локализованное и прошедшее ранее необходимую предобработку изображение лица.

Рассмотрим подробнее процесс, представленный на рис.1.

Изображение выглядит как диаграмма, текст, линия, чек

Автоматически созданное описание

*Рис. 1. Функциональная модель IDEF (уровень 0)*

На рис. 2 представлена IDEF 1 уровня декомпозиции.

**

*Рис. 2. Диаграмма декомпозиций IDEF (уровень 1)*

**Обзор существующих методов распознавания эмоций**

В процессе методологии принятия решений при распознавании образов выделяются три основных направления: *эвристические методы*, *математические методы* и *лингвистические (синтаксические) методы*.

*Эвристические методы*, такие как персептрон, базируются на опыте и интуиции разработчика системы. Их основным недостатком является отсутствие гарантии успешного решения задачи, что может привести к неопределенности при неудачных попытках применить интуитивные соображения.

*Математические методы* опираются на использование классического математического аппарата, включая методы линейного программирования, корреляционного анализа, и т.д. Среди них выделяются методы решающих функций, статистический метод, метод расстояния Хаусдорфа, метод взаимного отклонения и метод расстояния Хэмминга.

*Лингвистические (синтаксические, структурные) методы* применяются в случаях, когда образ представляет собой структуру из непроизводных элементов и признаков, описывающих связи между ними. Здесь широко используется аппарат алгебры логики и теории формальных языков.

В контексте методов распознавания эмоций рассмотрим классический способ определения эмоции по изображению лица, основанный на классификации ключевых точек, координаты которых можно получить, используя различные алгоритмы PDM, CML, AAM и другие. На изображении лица размечается множество точек, привязанных к положению бровей, глаз, губ, носа и челюсти, что позволяет осуществить частичный захват мимики. Нормализованные координаты точек подаются в классификатор (например, Random Forest) и получается базовое решение. Естественно положение лиц при этом должно быть выровнено. Однако даже при выполнении всех условий достигаемая точность будет отставать от более современных решений, поэтому данный метод можно назвать в определенном смысле устаревшим.

Также рассмотрим другой метод, основанный на геометрических характеристиках лица — это метод, предложенный Полом Экманом в их классической работе "Распознавание лицевых экспрессий" (позже FACS). Этот метод использует множество ключевых точек на лице, которые соответствуют различным мимическим мышцам. На основе положения этих точек и вычисления различных геометрических параметров, таких как расстояния, углы и площади, алгоритм может классифицировать основные эмоциональные состояния, которых, согласно идеологу данного подхода существует семь. Одной из основных проблем данного метода является изначальная его ориентированность на применение человеком, от того компьютерный расчет двигательных единиц может быть весьма затруднителен. Так, наличие препятствий в виде макияжа, татуировок, необычной прически и других может помешать поиску необходимых точек, а следовательно, и процессу распознавания.

Еще одним методом является метод главных компонент, который представляет изображения лиц в виде набора главных компонент. В задачах распознавания эмоций этот метод может использоваться для уменьшения размерности данных, без значительной потери информации об эмоциональных характеристиках, что позволяет сократить вычислительные затраты, тем самым влияя на быстродействие решения в целом.

*Сверточная нейронная сеть* использует особенности зрительной коры и обладает высокой точностью распознавания, которая может достигать значений более 90%. В числе прочих плюсов данного подхода стоит отметить отсутствие человеческого фактора при извлечении признаков., что отличает его от более традиционных подходов, где зачастую применяется ручная разработка признаков. Что важно, свёрточные нейронные сети обладают способностями к моделированию пространственных взаимосвязей и извлечению глубинных зависимостей, что дает методу возможность работы с глубинными знаниями и особенностями изображений, за счёт чего его результаты зачастую оказываются выше тех, что могут быть получены другими методами. Однако ее реализация затруднена, и существует риск переобучения при добавлении новых эталонных данных в базу.

**Поиск исходных данных для разработки решения**

Существуют несколько широко используемых наборов данных, содержащих изображения лиц людей с различными эмоциями. Один из наиболее известных - "FER2013" (The Facial Expression Recognition 2013), включает более 35 тысяч изображений лиц размером 48 на 48 пикселей в оттенках серого, разделенных на семь классов эмоций: радость, грусть, удивление, страх, отвращение, гнев и нейтральное состояние. Подготовка данных производилась таким образом чтобы лицо располагалось более или менее по центру и занимало примерно одинаковое, превышающее 50%, пространство на каждом изображении. Этот набор данных широко применяется для обучения и тестирования моделей распознавания эмоций.

Другой популярный набор данных "CK+" (Extended Cohn-Kanade) включает изображения лиц с различными эмоциями и информацию о движении мышц лица, способствуя более глубокому пониманию взаимосвязи между эмоциями и физиологическими процессами в лице.

Тем не менее, несмотря на наличие этих наборов данных, остаются незакрытые потребности в повышении качества обучающих наборов для распознавания эмоций. Эти потребности включают в себя несбалансированность классов эмоций, разнообразие культурных контекстов и выражений лиц, а также необходимость учета динамики эмоций во времени.

С развитием методов и подходов, использование упомянутых наборов данных и их развитие становится все более распространенной задачей в научных и практических исследованиях.

**Описание разработанного решения**

Благодаря проведенному анализу был выбран подход, основанный на использовании сверточной нейронной сети, для разработки собственного решения, поскольку данный нейросетевой метод эффективен в решении слабоструктурированных задач. Основным преимуществом выбранного метода является способность моделировать сложные нелинейные зависимости между входными и выходными переменными без явного установления точной формульной зависимости.

В качестве датасета для обучения предлагаемой модели был выбран "FER2013" (The Facial Expression Recognition 2013), несмотря на невысокую величину размера датасета для обучения сверточных нейронных сетей (приблизительно 35 тысяч изображений) и описанную несбалансированность классов эмоций. На данный момент этот датасет представляет собой наиболее конкурентоспособное решение по сравнению с аналогичными доступными наборами данных.

Архитектура сети представлена на рис. 3 Разработка собственной сети позволила решить определенные проблемы, в частности, обеспечить наличие лица на изображении при тестировании во всех случаях. Это означает, что при непосредственных испытаниях программы не возникало ситуаций, когда изображение с лицом человека расценивалось как изображение без лица, и была предоставлена возможность измерения характеристик сети.

Разработанное решение обладает точностью распознавания, превышающей 68%, что сопоставимо с другими публично доступными решениями для данного датасета и временем распознавания не превышающем заданный порог в 1 секунду.

*Изображение выглядит как текст, меню, Шрифт, число

Автоматически созданное описание*

*Рис. 3. Архитектура собственной сети*

**Сравнительная оценка решения относительно существующих разработок**

Полученные результаты требуется оценить, для этого прибегнем к методу, который предполагает сравнение с сопоставимыми аналогами. В качестве объекта сравнения выбрана исследовательская работа с аналогичной тематикой, однако с отличающимся методом разработки нейронной сети [4]. В данном случае, для проведения сравнительного анализа использованы результаты исследования, где применялась нейронная сеть с переносом обучения. Это подразумевает использование предварительно обученной модели ResNet50v2 (приблизительно 50 миллионов изображений) с измененным выходным классификатором, обученным на используемом в нашем исследовании датасете.

В рамках данного исследования, после прохождения 112 эпох точность работы модели достигла 73%. Результат данного эксперимента, хотя и высокий, сопровождается вероятностью переобучения модели. Под переобучением понимается способность модели, особенно при малом количестве обучающих данных, «запоминать» их, что позволяет ей показывать лучшие результаты на этих данных, однако это приводит к ухудшению способностей модели к обобщению, из-за чего реальная предсказательная сила модели может оказаться ниже полученной при обучении точности. Так как в описанном исследовании не проведена дополнительная проверка результатов на неиспользованных в процессе обучения изображениях, для частоты эксперимента ориентируемся на результаты, полученные моделью к моменту 35–40 эпох, с соответствующим уровнем точности в пределах 67-69%. Данный промежуток был выбран так как на нем начинается видимое расхождение между обучающей и валидационной ошибкой, где обучающая ошибка продолжает сокращаться, а валидационная выходит на «плато» с разнонаправленными колебаниями показателя, что указывает на возникновение на этом этапе переобучения. Взятый для оценки результат, соответствующий описанному эпохальному периоду - согласуется с нашими результатами.

Также было проведено сравнение с наиболее популярными решениями, представленными на форуме Kaggle. Показатели, достигнутые в этих решениях, либо ниже, либо условно сопоставимы с достигнутыми нашей моделью показателями. Это позволяет говорить о конкурентоспособности разработанной модели.

Изображение выглядит как текст, линия, диаграмма, График

Автоматически созданное описание

*Рис. 4. Изменение метрик «accuracy» и «loss» на обучающем и валидационном наборах данных в процессе обучения нейронной сети, построенной на основе ResNet50*

Алгоритм полного распознавания, начиная от изображения и завершая выявлением эмоций, разработан с использованием инструментария библиотеки OpenCV для языка программирования Python 3. Средствами данной библиотеки осуществляется поиск лиц на входящих фотографиях и их предварительная обработка с соответствием стандартам нашей модели. Эта обработка включает в себя выделение квадратной области, содержащей лицо, перевод изображения в черно-белую гамму и изменение размера до 48 на 48 пикселей, что соответствует входным параметрам нашей модели. Графическое представление полного алгоритма распознавания изображено на рис. 5.

*Изображение выглядит как текст, черно-белый, Симметрия, снимок экрана

Автоматически созданное описание*

*Рис. 5. Блок-схема работы алгоритма распознавания*

**Заключение**

В рамках проведенного исследования были рассмотрены существующие подходы к решению задачи распознавания эмоций и проанализированы датасеты, подходящие для этого. Было разработано программное решение воплощающее в себе алгоритм распознавания полного цикла, от поступления на вход неподготовленного изображения, до, в качестве выхода, определения класса эмоции для каждого представленного на изображении лица. В своей специфике оно основано на применении свёрточной нейронной сети собственной архитектуры. Исследование имеет перспективы для применения в различных сферах, от использования в качестве научной базы будущих работ в области, до, непосредственного применения в процессах диагностики психических расстройств в медицине и улучшения стратегий в маркетинге и рекламе.

**Список литературы**

1. Viola, P. Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. –Kauai, Hawaii, 2008. – V. 1. – p. 513–520.
2. Viola, P. Robust real-time face detection // International Journal of Computer Vision. – 2005. –V. 58. – № 2. – p. 137–155.
3. Метод Виолы-Джонса (Viola-Jones) как основа для распознавания лиц. URL: https://habrahabr.ru/post/133826/ (Дата обращения 02.02.2024)
4. Галичий Д.А., Афанасьев Г.И., Нестеров Ю.Г. Распознавание эмоций человека при помощи современных методов глубокого обучения // E-Scio. 2021. №5 (56). URL: https://cyberleninka.ru/article/n/raspoznavanie-emotsiy-cheloveka-pri-pomoschi-sovremennyh-metodov-glubokogo-obucheniya (дата обращения: 30.01.2024).