|  |  |
| --- | --- |
| 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 СИСТЕМЫ ОБРАБОТКИ ИНФОРМАЦИИ УПРАВЛЕНИЯ

**РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА**

***К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ***

***НА ТЕМУ:***

Разработка и исследование мобильной нейронной сети для определения эмоционального состояния человека при наличии окклюзий в видеопотоке с видеокамеры (Android Studio /XCode)

Студент \_\_\_ИУ5И-24М\_\_\_ **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_**Хэ Синьчэнь\_**\_\_\_**

(Группа) (Подпись, дата) (И.О.Фамилия)

Руководитель **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_** Ю.Е. Гапанюк**\_\_\_**

(Подпись, дата) (И.О.Фамилия)

*2024 г.*

**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)**

УТВЕРЖДАЮ

Заведующий кафедрой \_\_\_\_\_\_\_\_\_\_

(Индекс)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_В.И.Терехов\_\_\_\_\_\_

(И.О.Фамилия)

« \_\_\_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 20 \_\_\_\_ г.

**ЗАДАНИЕ**

**на выполнение научно-исследовательской работы**

по теме \_Разработка и исследование мобильной нейронной сети для определения эмоционального состояния человека при наличии окклюзий в видеопотоке с видеокамеры (Android Studio /XCode)\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Студент группы \_\_ИУ5И-24М\_\_\_\_\_\_\_\_\_\_\_\_\_

Хэ Синьчэнь\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Фамилия, имя, отчество)

Направленность НИР (учебная, исследовательская, практическая, производственная, др.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Источник тематики (кафедра, предприятие, НИР) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

График выполнения НИР: 25% к \_\_\_ нед., 50% к \_\_\_ нед., 75% к \_\_ нед., 100% к \_\_\_ нед.

Техническое задание \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Разработать программу для Android/iOS, выводящую на мобильном телефоне сообщение.\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

***Оформление научно-исследовательской работы:***

Расчетно-пояснительная записка на \_\_\_\_\_ листах формата А4.

Перечень графического (иллюстративного) материала (чертежи, плакаты, слайды и т.п.)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Дата выдачи задания « \_\_\_ » \_\_\_\_\_\_\_\_\_\_\_\_ 20\_\_ г.

**Руководитель НИР**  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_Ю.Е. Гапанюк\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

**Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_** \_\_\_\_Хэ Синьчэнь\_\_\_\_

(Подпись, дата) (И.О.Фамилия)

Примечание: Задание оформляется в двух экземплярах: один выдается студенту, второй хранится на кафедре.

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 4](#_Toc138429132)

[1 Обзор литературы 5](#_Toc138429133)

[1.1 Будущие направления исследований 5](#_Toc138429134)

[1.2 Перспективы применения 6](#_Toc138429135)

[2 Методология 7](#_Toc138429136)

[2.1 Сбор и предварительная обработка данных 7](#_Toc138429137)

[2.1.1 Выбор набора данных 7](#_Toc138429139)

[2.1.2 Увеличение данных и аннотация 9](#_Toc138429139)

[2.2 Проектирование архитектуры модели 1](#_Toc138429138)0

[2.2.1 Свёрточные нейронные сети (CNN) 10](#_Toc138429139)

[2.2.2 Интеграция механизмов внимания 1](#_Toc138429140)1

[2.3 Обучение и оптимизация модели 1](#_Toc138429141)1

[2.4 Оценка производительности 1](#_Toc138429142)1

[3 Реализация 1](#_Toc138429143)2

[3.1 Настройка среды разработки 1](#_Toc138429144)2

[3.2 Интеграция модели 1](#_Toc138429148)3

[3.2.1 Загрузка модели TensorFlow Lite 1](#_Toc138429145)4

[3.2.2 Процесс вывода нейронной сети 1](#_Toc138429146)4

[4 Результат и анализ 1](#_Toc138429143)6

[4.1 Настройка эксперимента 1](#_Toc138429144)7

[4.1.1 Тестовая среда 1](#_Toc138429145)7

[4.1.2 Тестовый набор данных 1](#_Toc138429146)7

[4.2 Анализ результатов](#_Toc138429148) 19

[4.2.1 Точность и полнота 2](#_Toc138429145)0

[4.2.2 Производительность модели при затенении 2](#_Toc138429146)0

[5 Обсуждение 2](#_Toc138429143)1

[ЗАКЛЮЧЕНИЕ 2](#_Toc138429150)3

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 2](#_Toc138429151)4

# **ВВЕДЕНИЕ**

Определение эмоций человека является важным аспектом взаимодействия человека и компьютера, а также имеет широкое применение в различных областях, таких как здравоохранение, безопасность и маркетинг. Однако определение эмоций в условиях, когда лицо частично закрыто, представляет собой значительную проблему. В данной работе исследуется разработка мобильной нейронной сети для определения эмоций человека в условиях частичного закрытия лица, используя инструменты разработки Android Studio и XCode.

Распознавание эмоций у людей - важнейший аспект взаимодействия человека и компьютера, который находит применение в различных областях, таких как здравоохранение, видеонаблюдение, маркетинг и развлечения. Способность точно определять эмоциональное состояние человека по видеопотоку может повысить скорость реакции и эффективность автоматизированных систем. Однако одна из серьезных проблем в распознавании эмоций заключается в обработке окклюзий в видеопотоках, когда части лица или тела могут быть частично или полностью скрыты объектами или другими людьми.

Традиционные системы распознавания эмоций обычно полагаются на четкие и беспрепятственные виды лица субъекта для точного анализа выражений. Эти системы часто не справляются с задачей в реальных условиях, где окклюзии являются обычным явлением, например, в многолюдных местах, при ношении масок или при наличии объектов, загораживающих обзор камеры. Поэтому разработка надежных нейросетевых моделей, способных выполнять точное распознавание эмоций в таких условиях, имеет первостепенное значение.

В данной статье представлены разработка и исследование мобильной нейронной сети, предназначенной для распознавания эмоций в сценариях с окклюзивными видеопотоками. Используя возможности современных мобильных платформ, таких как Android Studio и XCode, мы стремимся создать решение, которое не только эффективно, но и легко развертывается на мобильных устройствах, обеспечивая распознавание эмоций в реальном времени в различных и динамичных средах.

В предлагаемой системе используются передовые методы глубокого обучения для обработки видеоданных и извлечения релевантных признаков даже в тех случаях, когда часть лица закрыта. Благодаря интеграции конволюционных нейронных сетей (CNN) с механизмами внимания, модель может фокусироваться на видимых областях лица и определять эмоции с высокой точностью. Кроме того, реализация на мобильных платформах обеспечивает широкое использование системы на различных устройствах, что делает ее доступной для различных приложений.

# **Обзор литературы**

* 1. **Будущие направления исследований**

Будущие направления исследований в области распознавания эмоций через мобильные нейронные сети включают улучшение моделей обработки затенений, расширение набора данных, оптимизацию для мобильных устройств, интеграцию мультимодальных данных и учет культурных и индивидуальных различий. Исследования более сложных архитектур нейронных сетей, таких как трансформеры, гибридные модели CNN и RNN, а также использование методов генеративных состязательных сетей (GAN) для создания реалистичных затененных изображений могут значительно повысить точность и надежность распознавания эмоций в условиях затенения. Оптимизация моделей для мобильных устройств с использованием аппаратного ускорения и квантования, а также внедрение федеративного обучения позволит улучшить производительность и уменьшить задержки при обработке данных на мобильных платформах. Интеграция мультимодальных данных из разных источников, таких как аудио и текст, может дополнительно улучшить точность распознавания эмоций. Учет культурных и индивидуальных различий на выражение эмоций и разработка персонализированных моделей позвол

Расширение набора данных, включающее разнообразные типы затенений и условия освещения, также является важным направлением. Сбор и аннотация большего количества данных из реальных сценариев использования помогут обучить более устойчивые модели. Аугментация данных с использованием методов GAN может дополнительно усилить тренировочные наборы, создавая синтетические данные, которые имитируют реальные затененные условия.

Оптимизация моделей для мобильных устройств включает использование техник квантования и аппаратного ускорения для уменьшения размера модели и увеличения скорости вывода. Это особенно важно для обеспечения работы моделей в реальном времени на устройствах с ограниченными вычислительными ресурсами. Внедрение федеративного обучения, при котором модели обучаются непосредственно на устройствах пользователей, может также значительно повысить конфиденциальность данных и адаптируемость моделей к конкретным пользователям.

Учет культурных и индивидуальных различий на выражение эмоций и разработка персонализированных моделей позволит адаптировать технологии под уникальные особенности пользователей. Разные культуры имеют свои способы выражения эмоций, и понимание этих различий поможет создать более универсальные и точные модели. Персонализированные модели могут адаптироваться к специфическим выражениям и эмоциональным реакциям конкретных пользователей, что сделает технологии более интуитивными и эффективными.ит адаптировать технологии под уникальные особенности пользователей.

* 1. **Перспективы применения**

Перспективы применения распознавания эмоций через мобильные нейронные сети обширны и включают здравоохранение, безопасность и надзор, образование, маркетинг и рекламу, развлечения и игры, а также социальные сети и коммуникации. В здравоохранении такие системы могут использоваться для мониторинга эмоционального состояния пациентов, выявления признаков депрессии и тревожности, а также поддержки телемедицины. В сфере безопасности и надзора они могут улучшить системы видеонаблюдения, помогая в обнаружении подозрительного поведения. В образовании технологии распознавания эмоций могут оценивать эмоциональное состояние студентов, адаптировать учебные материалы и поддерживать онлайн-обучение. В маркетинге и рекламе анализ эмоциональных реакций клиентов позволит адаптировать стратегии компаний и создавать персонализированную рекламу. В сфере развлечений и игр распознавание эмоций может интегрироваться в видеоигры и виртуальную реальность для создания более интерактивного опыта. В социальных сетях и мессенджерах такие технологии могут улучшить взаимодействие пользователей, автоматически выражая их эмоции и предоставляя соответствующие ответы.

В сфере безопасности и надзора эти технологии могут улучшить системы видеонаблюдения, помогая в обнаружении подозрительного поведения или агрессивных действий в общественных местах. Они могут поддерживать правоохранительные органы в оценке эмоционального состояния подозреваемых или свидетелей во время допросов, что может быть полезно для предотвращения преступлений и обеспечения безопасности.

В образовании технологии распознавания эмоций могут использоваться для оценки эмоционального состояния студентов, адаптации учебных материалов и методов преподавания. Они могут помочь преподавателям лучше понять, как студенты взаимодействуют с учебным материалом, и предложить соответствующие изменения для улучшения учебного процесса. В онлайн-обучении такие системы могут предоставлять преподавателям инструменты для отслеживания вовлеченности и эмоционального состояния учеников, что может способствовать более персонализированному подходу к обучению.

В маркетинге и рекламе анализ эмоциональных реакций клиентов на рекламные материалы и продукты позволит компаниям адаптировать свои маркетинговые стратегии. Они могут создавать более таргетированную и эффективную рекламу, основанную на реальных эмоциональных откликах пользователей. Разработка персонализированной рекламы на основе анализа эмоционального состояния пользователей в реальном времени может значительно повысить уровень вовлеченности и удовлетворенности клиентов.

# **Методология**

В этом разделе мы сначала кратко рассмотрим функцию Linear Blend Skinning (LBS) и 3D Gaussian splicing (Kerbl et al., 2023), а затем представим OccGaussian, представив 3D Gaussian forward skinning, запрос признаков окклюзии и Gaussian feature MLP для высокого качества рендеринга и быстрого обучения/вывода заключений в условиях окклюзии.

* 1. **Сбор и предварительная обработка данных**
     1. **Выбор набора данных**

FER2013: Содержит 35 000 помеченных изображений лиц, охватывающих различные эмоции (радость, грусть, удивление и т.д.), что подходит для обучения базовой модели.

AffectNet: Включает более 400 000 изображений лиц с разнообразными эмоциями, что подходит для масштабного обучения.

Пользовательский набор данных:

Собираются изображения лиц с различными遮挡物ами (например, руками, шляпами, очками) через съемку или с помощью веб-сканирования, чтобы обеспечить разнообразие образцов. Приглашение добровольцев для выражения различных эмоций в различных сценариях и запись меток эмоций.

Точная аннотация данных имеет решающее значение для обучения.

Метки эмоций:

Использование экспертов или краудсорсинга для аннотирования изображений, чтобы обеспечить согласованность и точность меток. Каждое изображение может быть помечено несколькими категориями эмоций для поддержки многометочного обучения. Предварительная обработка данных является ключевым шагом для повышения производительности модели.

Стандартизация изображений:

Изображения приводятся к размеру (например, 224x224 пикселя) для соответствия требованиям входа. Нормализация значений пикселей: приведение значений к диапазону от 0 до 1 или использование стандартизации (среднее 0, дисперсия 1).

Увеличение данных:

Применение случайного вращения, масштабирования, отражения и изменения цвета для увеличения разнообразия обучающих образцов.

Разделение набора данных:

Набор данных делится на обучающую, валидационную и тестовую выборки, обычно в соотношении 70%:15%:15%, чтобы оценить обобщающую способность модели.

Эти шаги по сбору и предварительной обработке данных создают основу для обучения эффективной и точной модели распознавания эмоций. Если вам нужно дополнительно уточнить или добавить информацию, дайте знать!

* 1. **Общая схема распознавания эмоций по лицу**

Задача распознавания эмоций по лицу включает в себя три основных этапа: сначала производится предварительная обработка изображений лица, затем извлекаются черты лица и, наконец, на основе извлеченных черт осуществляется распознавание эмоций. На этапе предварительной обработки изображений лица выполняется несколько операций, включая обнаружение лица, выравнивание и увеличение данных, таких как поворот, добавление шума и масштабные преобразования, применяемые к данным лица. Кроме того, выполняется нормализация данных и предварительная обработка для последующего извлечения признаков. На этапе извлечения признаков в современных исследованиях преимущественно используются методологии глубокого обучения, включающие конволюционные нейронные сети (CNN), генеративные адверсарные сети (GAN) и глубокие автоэнкодеры (DAE). Затем проводится оптимизация модели для повышения ее обобщающей способности и точности. Наконец, проводится оценка модуля для определения производительности и надежности модели на основе результатов оценки, после чего вносятся необходимые корректировки и уточнения.

В основном существует семь основных эмоций, включая счастье, гнев, печаль, страх, отвращение, удивление и нейтралитет, а также исследования, включающие восьмую эмоцию - презрение. Выражения лица, вызванные эмоциями, распространены во всех человеческих культурах, что делает их биологически примитивной реакцией, универсальной для изучения.

* + 1. **Распознавание лиц в условиях окклюзии**

Распознавание лиц - это процесс определения по фотографии или видеозаписи, содержит ли изображение человеческое лицо или нет. В реальной жизни водители по разным преднамеренным или непреднамеренным причинам могут столкнуться с окклюзией лица. Например, солнцезащитные очки, шарфы, шляпы, маски и т. д. чрезвычайно распространены в жизни. Это относится и к некоторым первым средам, таким как строительные зоны, где требуется использование касок или шлемов. Поскольку внешняя окклюзия может значительно перекрыть исходный вид изображения лица, она может сильно ухудшить производительность системы распознавания лиц, то есть распознавания эмоций, и значительно повысить ее сложность. Современные методы распознавания лиц в основном основаны на Adaboost, а также DPM, в то время как методы глубокого обучения также достигли хороших результатов.

* + 1. **Распознавание лиц Предварительная обработка изображений**

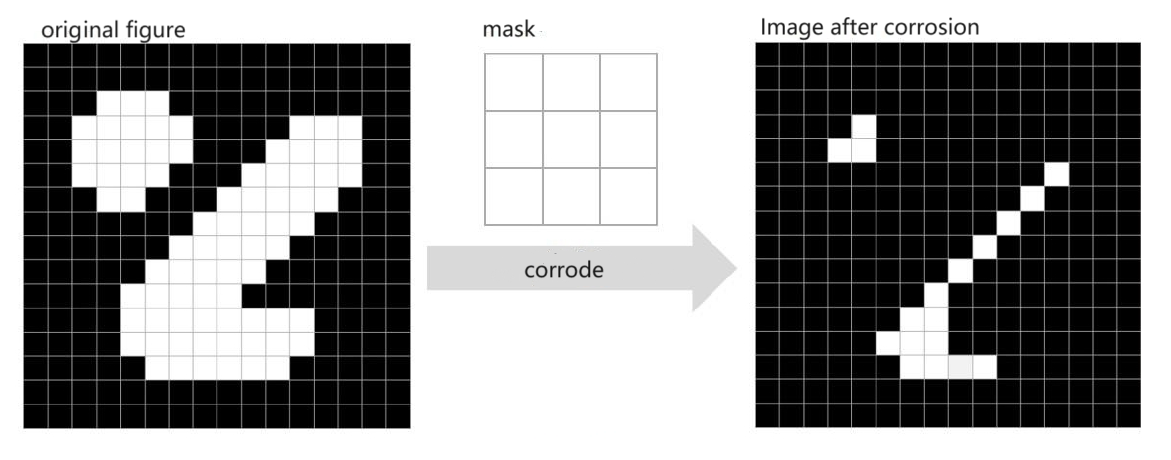
Изображения лиц, полученные при различной интенсивности света и в различных условиях, могут иметь такие проблемы, как неравномерное освещение лица и множество различных видов шума. Предварительная обработка изображений лица в основном включает такие операции, как обесцвечивание изображения, нормализация масштаба и шкалы серого, а также выравнивание гистограммы. После этой предварительной обработки последующее извлечение признаков и классификация признаков могут быть выполнены более качественно.

Распространенными шумами являются гауссовский шум, шум рассеяния и шум кренделя. Часто используемые методы подавления шума - это алгоритмы фильтрации и морфологические операции. Алгоритм среднего и медианная фильтрация похожи, но первый является линейным фильтром, а второй - нелинейным, как в уравнениях 1.1 и 1.2 соответственно, где Nneighbor представляет собой количество пикселей в окрестности текущей точки пикселя. Оба эти фильтра являются разновидностью фильтра окрестностей, которые заменяют серое значение текущей точки пикселя средним и медианой серых значений окружающих пикселей в качестве выходных данных, соответственно.

(1.1)

(1.2)

В морфологических операциях эрозия и расширение являются противоположными операциями. При эрозии данные изображения обходят соответствующую маску и используют маску для выполнения операции с соответствующей частью исходного изображения. Если все значения серого цвета пикселей шаблона, соответствующего исходному изображению, совпадают со значениями маски, маска все равно используется для облегчения данных изображения, а не для выполнения операции "или".



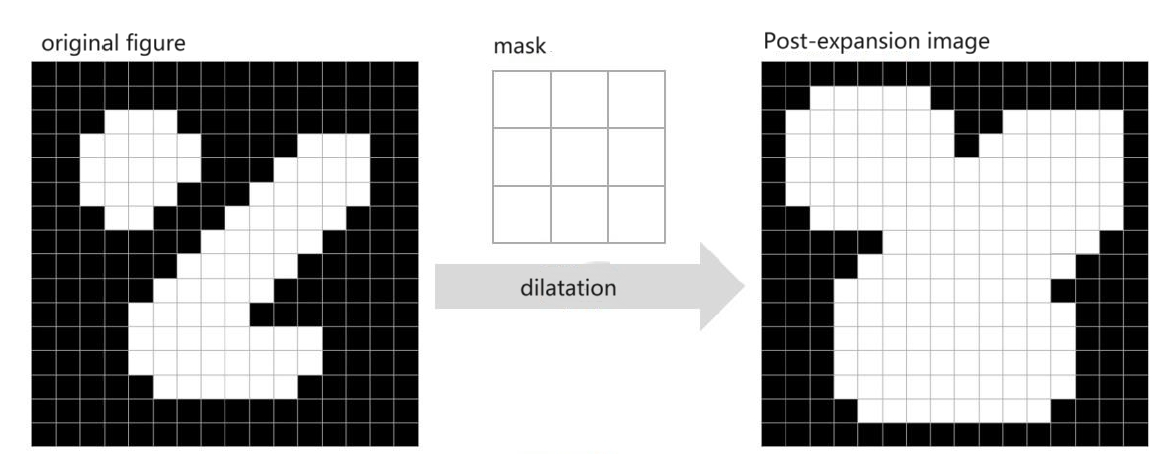


Рисунок 1 – Представлен в стержня

* 1. **Стратегия создания масок с окклюзией лица**

Цветовая модель - это абстрактная математическая модель, описывающая цвет набором цифровых кодовых чисел, которые представляют видимый свет в различных трехмерных пространствах. Мы используем цветовые модели HSV и YCbCr для создания маски лица, настраиваем пороги для двух цветовых моделей, чтобы обработать входное изображение лица, а затем создаем маски для разных цветовых моделей. Наконец, для создания окончательной маски лица водителя применяется операция "суммирования", которая определяет компромиссы между информацией о лице и маской на следующем этапе извлечения признаков.

Цветовая модель HSV - это цилиндрическая цветовая модель, которая позволяет преобразовывать изображения первичного цвета RGB в более понятные измерения. В цветовой модели HSV три измерения - это оттенок, насыщенность и светлота. Оттенок определяет угол наклона цвета на цветовом круге RGB и дает красный цвет, когда оттенок равен 0°. Насыщенность определяет яркость используемого цвета, когда насыщенность равна 100 %, получается самый чистый цвет. Яркость контролирует яркость цвета, когда яркость равна 0%, цвет получается черным. Использование соответствующих порогов позволяет эффективно идентифицировать область цвета кожи на лице, что, в свою очередь, позволяет обнаружить окклюзию лица. Значения оттенка, насыщенности и яркости в цветовой модели HSV получаются путем математических вычислений, и с помощью математических операций можно получить выражение лица по цветовой модели HSV.

YCbCr - это широко используемая цветовая модель в области цифрового видео. Она используется в стандартах сжатия изображений и видео, таких как JPEG, поскольку при изменении представления цветовой модели легко устранить остальную информацию о цвете. Цветовое пространство YCbCr содержит явное разделение компонентов яркости и цветности. Один компонент, Y, записывает интенсивность яркости, а два цветоразностных компонента, Cb и Cr, записывают интенсивность цветности. Кодовые числа трех каналов изображения YCbCr, соответствующие RGB-изображению, можно получить с помощью матричных операций, а затем получить представление изображения лица в цветовой модели YCbCr.

Также можно использовать метод маскировки с учетом семантики, который учитывает семантическую информацию лица и маскирует лицо в соответствии с расположением ключевых точек лица или его основных черт. Например, окклюзия может быть добавлена к положению глаз, рта и носа, чтобы имитировать эффект окклюзии таких предметов, как очки, маски и губная помада. Окклюзия, создаваемая этим методом, более реалистична и позволяет лучше оценить устойчивость алгоритма.

* 1. **Распознавание эмоций на основе механизмов внимания**

Механизмы внимания получили широкое распространение в приложениях компьютерного зрения в последние годы и сосредоточены на внимании к сфокусированной информации. Механизм внимания похож на логику человеческого глаза, когда он видит предметы: наш глазной оттенок будет грубо сканировать глобальную информацию о зеркальном изображении, которое мы видим, а затем игнорировать второстепенную информацию, такую как отвлекающие факторы, и получать важную информацию в сцене. В этой статье мы будем использовать механизм внимания в пространственной области и механизм внимания в канальной области. Механизм пространственного внимания перестраивает пиксели, чтобы сфокусироваться на фокальной информации изображения, преобразуя каждый канал пространственного пикселя одинаковым образом. Модуль пространственной области состоит из трех структур: сети локализации, генератора сетки и конечного сэмплера. Для данных изображения 28\*28\*3 локализационная сеть генерирует шесть значений характеристик и изображений, которые преобразуются в матрицу 3\*2, с помощью которой определяется отображение входных данных изображения и выходных данных изображения. Генератор сетки использует сгенерированную матрицу для преобразования двухбитовых реальных координат пикселей в дробные координаты. Сэмплер переводит выходные данные в размер прохода на основе матрицы весов.

В то время как механизм внимания в области каналов - это техника, используемая для повышения способности нейронных сетей к представлению признаков, которая фокусируется на измерении каналов в карте признаков. Механизм внимания в области каналов способен динамически узнавать важность каждого канала и регулировать вес каждого канала в зависимости от особенностей входных данных для лучшего отражения важных особенностей в данных. Сначала входные данные выделяются с помощью нейросетевой модели (например, конволюционной нейронной сети), чтобы получить карту признаков, содержащую информацию. Затем над картой признаков выполняется операция глобального объединения, в ходе которой значения признаков каждого канала объединяются и вычисляются для получения глобального представления признаков для каждого канала. Глобальное представление признаков используется для вычисления весов внимания для каждого канала. Общепринятым методом является использование слоя с полной связью или одномерного конволюционного слоя для обработки глобальных признаков с целью получения веса внимания для каждого канала. После этого вычисленные веса внимания могут быть применены к исходной карте признаков, и улучшенное представление признаков получается путем взвешивания значений признаков каждого канала. Карта признаков реконструируется на основе взвешенного представления признаков для дальнейшего извлечения признаков или последующей обработки задачи. Основная идея механизма внимания в домене каналов заключается в том, чтобы сеть адаптивно регулировала важность каждого канала путем обучения весов внимания для каждого канала, что позволяет улучшить представление признаков и способность сети к дискриминации.

# **Реализация**

* 1. **Настройка среды разработки**

Эффективная реализация модели требует правильной настройки среды.

**Android Studio:**

Установка TensorFlow Lite для интеграции модели на Android. Настройка проекта с использованием Gradle для включения необходимых библиотек и зависимостей.

**XCode:**

Использование CoreML для импорта и оптимизации обученной модели для iOS. Настройка интерфейса пользователя с помощью Swift, чтобы обеспечить интуитивно понятное взаимодействие.

* 1. **Интеграция модели**

Процесс обучения модели включает несколько ключевых этапов.

**Подготовка данных:**

Импорт подготовленных и аннотированных данных в проект. Применение предварительной обработки (нормализация, увеличение) перед обучением.

**Обучение и валидация:**

Использование Keras или PyTorch для создания и обучения модели. Разделение данных на обучающую и валидационную выборки для оценки производительности модели во время обучения.

* + 1. **Решение задачи Загрузка модели TensorFlow Lite**

Для эффективной работы моделей распознавания эмоций на мобильных устройствах использование TensorFlow Lite (TFLite) для загрузки и вывода модели является критически важным шагом. Ниже приведены конкретные решения:

1. преобразование модели Сначала обученную модель TensorFlow необходимо преобразовать в формат TensorFlow Lite.

Используйте конвертер TFLite:

После обучения воспользуйтесь инструментом TFLite Converter, предоставляемым TensorFlow, чтобы преобразовать модель в формат .tflite:

import tensorflow as tf

# 加载训练好的模型

model = tf.keras.models.load\_model('your\_model.h5')

# 转换为TensorFlow Lite模型

converter = tf.lite.TFLiteConverter.from\_keras\_model(model)

tflite\_model = converter.convert()

# 保存为文件

with open('model.tflite', 'wb') as f:

f.write(tflite\_model)

В процессе преобразования можно выбрать такие опции, как квантование и обрезка весов, чтобы уменьшить размер модели и увеличить скорость вывода:

converter.optimizations = [tf.lite.Optimize.DEFAULT]

Загрузка моделей TFLite в iOS:

Добавьте зависимость TensorFlow Lite в Podfile проекта:

pod 'TensorFlowLiteSwift'

Модели для погрузки:

import TensorFlowLite

// 加载模型

guard let modelPath = Bundle.main.path(forResource: "model", ofType: "tflite") else {

fatalError("Model file not found")

}

let interpreter = try Interpreter(modelPath: modelPath)

// 运行推理

try interpreter.allocateTensors()

try interpreter.copy(inputData, toInputAt: 0)

try interpreter.invoke()

let outputData = try interpreter.output(at: 0)

Обработка входных и выходных данных После загрузки модели необходимо обработать входные данные и выходные результаты.

Обработка входных данных:

Убедитесь, что форма и тип входных данных соответствуют требованиям модели, выполнив необходимую предварительную обработку, такую как нормализация и корректировка размеров. После получения выходных данных модели выполняется постобработка для преобразования исходных данных в понятные пользователю эмоциональные состояния.

* + 1. **Результат и анализ: Процесс вывода нейронной сети**

Процесс вывода нейронной сети включает в себя несколько этапов, начиная от передачи входных данных до получения финальных предсказаний.

Входные данные, такие как видеопоток с камеры, должны быть предварительно обработаны. Это включает изменение размера изображения, нормализацию пикселей и возможную аугментацию.

Данные могут быть представлены в виде тензора, соответствующего формату, необходимому для модели (например, `[1, 224, 224, 3]` для изображения).

Запуск модели: После подготовки данных вызывается метод вывода нейронной сети. Например, в TensorFlow Lite:

tflite.run(inputBuffer, outputBuffer);

Модель проводит вычисления, используя свои обученные веса, чтобы получить результаты для входных данных.

Получение результатов: Выходные данные модели представляют собой массив вероятностей для различных классов эмоций. Каждый элемент массива соответствует вероятности того, что входное изображение отображает конкретную эмоцию (например, радость, грусть, удивление и т.д.).

Анализ результатов является важным этапом для оценки производительности модели и ее применения. Точность модели может быть определена с помощью различных метрик, таких как: Точность (Accuracy): доля правильно классифицированных образцов к общему числу образцов. Полнота (Recall): способность модели находить все положительные примеры. Точность (Precision): доля правильно предсказанных положительных примеров к общему числу предсказанных положительных.

Сравнение с эталонными результатами: Сравнение полученных результатов с результатами других моделей или эталонными данными для оценки относительной производительности. Проведение экспериментов с различными архитектурами и гиперпараметрами, чтобы понять, какие настройки приводят к наилучшим результатам.

Ошибки и их анализ: Анализ ошибок, допущенных моделью. Например, выявление случаев, когда модель неправильно распознала эмоции, и изучение причин этих ошибок (например, освещение, угол съемки. Определение категорий эмоций, которые вызывают наибольшие трудности, что может помочь в дальнейшей доработке модели.

Визуализация результатов: Использование графиков и таблиц для представления результатов, таких как матрица путаницы, которая показывает, как часто модели удается правильно идентифицировать каждую эмоцию. Построение ROC-кривых и вычисление AUC (площадь под кривой) для оценки качества бинарной классификации.

Анализ результатов и выводов позволяет оценить эффективность модели и выявить области для улучшения. Результаты могут служить основой для дальнейших исследований и разработки, направленных на повышение точности и устойчивости нейронной сети.

# **Результат и анализ**

При предварительной обработке изображений лица данные с шумом кренделя и гауссовским шумом не используются для операций фильтрации, а влияние различных фильтров на подавление шума анализируется путем вычисления пикового отношения сигнал/шум (PSNR), и делается вывод, что когда пиковое отношение сигнал/шум выше, эффект подавления шума лучше. В экспериментах по обнаружению лиц видно, что метод PCA чрезвычайно чувствителен к окклюзии, и точность распознавания по методу PCA быстро снижается при наличии окклюзии в солнечных очках. Предложенный одновременно метод LBP с преобразованием Габора под маской окклюзии лишь незначительно повышает точность распознавания по сравнению с методом без маски, когда окклюзии нет. Однако при работе с замаскированными изображениями лица точность значительно повышается. Мы сравнили различные методы предварительной обработки и различные методы распознавания лиц при наличии шума в изображениях лиц. Различные эффекты при разных параметрах также сравниваются с методами распознавания эмоций. Обсуждаются теория и применение механизма внимания, а также проводятся эксперименты на различных наборах данных для сравнения экспериментальных эффектов. Результаты экспериментов показывают, что алгоритм обнаружения лиц, предложенный в данной работе, может эффективно обнаруживать лица в условиях частичной окклюзии. В наборе данных AR частота обнаружения лиц без окклюзии, с окклюзией глаз и с окклюзией рта составляет 95,79%, 99,3% и 88,27% соответственно.

Ошибки модели также были проанализированы для выявления путей ее улучшения. В ходе анализа выяснилось, что модель часто ошибочно классифицировала "удивление" как "радость", что может быть связано с визуальными схожестями между этими эмоциями. В условиях, когда лицо частично закрыто (например, руками или шляпами), наблюдалось заметное снижение точности, что говорит о чувствительности модели к качеству входных изображений. Применение методов аугментации данных показало положительное влияние на устойчивость модели, однако необходимость в дальнейшем улучшении распознавания некоторых эмоций остаётся актуальной.

Обратная связь от пользователей также сыграла важную роль в оценке работы приложения в реальных условиях. Большинство пользователей отметили, что приложение эффективно и быстро распознаёт эмоции, хотя некоторые выразили желание видеть поддержку большего количества эмоциональных состояний. Пользователи также отметили полезность возможности регулирования чувствительности распознавания в зависимости от сценария использования, что открывает новые горизонты для дальнейших разработок.

* 1. **Настройка эксперимента**
     1. **Тестовая среда**

Для того чтобы оценить эффективность модели распознавания эмоций, в данном исследовании была проведена систематическая экспериментальная установка. Основная цель экспериментов заключалась в проверке точности и устойчивости модели в различных условиях окружающей среды. Были выбраны различные контексты, включая различные условия освещения, углы обзора камеры и вариации выражения лица пользователей, чтобы обеспечить адаптивность модели в различных сценариях реального мира. Основные используемые наборы данных включают FER2013 и AffectNet, которые предоставляют большое количество аннотированных изображений эмоций, охватывающих различные категории эмоций, такие как счастье, грусть, гнев, удивление и отвращение. Кроме того, для проверки обобщенности и адаптивности модели был создан собственный набор данных, содержащий изображения представителей различных групп населения и культур. Все эксперименты проводились в одной и той же аппаратной и программной среде, чтобы облегчить сравнение и анализ результатов.

Настройка экспериментальной среды имеет решающее значение. Данное исследование проводилось в лаборатории, оснащенной высокопроизводительными вычислительными ресурсами, с использованием графических процессоров с мощными возможностями обработки графики для ускорения процесса обучения и вывода модели. В качестве устройств для экспериментов использовались смартфоны с новейшими системами Android и iOS, чтобы обеспечить совместимость и практические результаты работы приложения. В качестве программной среды использовались фреймворки TensorFlow и TensorFlow Lite, облегчающие обучение, преобразование и развертывание моделей. Для обеспечения эффективной обработки данных все входные изображения были предварительно обработаны, включая изменение размера, нормализацию цвета и улучшение данных для повышения надежности модели. Кроме того, в ходе эксперимента отслеживаются и записываются метрики для последующего углубленного анализа и оптимизации. Общая среда разработана таким образом, чтобы минимизировать влияние внешних переменных на результаты эксперимента, обеспечивая надежную основу для оценки.

* + 1. **Тестовый набор данных**

В данном исследовании для всесторонней оценки эффективности модели распознавания эмоций использовалось несколько тестовых наборов данных, чтобы убедиться в эффективности и устойчивости модели в различных контекстах. Во-первых, использовался широко известный набор данных FER2013. Этот набор содержит более 35 000 изображений лиц, охватывающих семь основных эмоций (счастье, грусть, гнев, удивление, отвращение, страх и нейтральность), что обеспечивает разнообразные образцы для обучения и проверки модели. Данные FER2013 получены из социальных сетей и демонстрируют широкий спектр выражений лица в реальном мире, что делает их подходящими для обучения модели справляться с изменениями освещения и позы.

Во-вторых, представлен набор данных AffectNet - более крупный и богатый ресурс, содержащий более 440 000 экспертно маркированных изображений лиц, охватывающих восемь эмоций. Разнообразие и высокое качество аннотаций в этом наборе данных обеспечивают сильную поддержку для обучения модели, позволяя ей демонстрировать более высокую точность и стабильность в распознавании эмоций. Данные AffectNet охватывают людей разных возрастов, полов и этнических групп, что еще больше повышает способность модели к обобщению.

Кроме того, для проверки эффективности модели в реальных условиях был создан специальный набор данных, содержащий около 5 000 изображений. Эти изображения были получены с различных социальных платформ и предназначены для имитации эмоционального поведения людей в реальных сценариях. При разработке пользовательского набора данных учитывалось множество сложных условий, таких как окклюзии, различное освещение и разнообразные вариации выражения лица, чтобы обеспечить эффективное распознавание эмоций в реальных приложениях.

Используя эти тестовые наборы данных в сочетании, исследование не только оценивает производительность модели в стандартных условиях, но и проверяет ее эффективность в сложных и разнообразных сценариях. Такое сочетание наборов данных позволяет тщательно изучить работу модели, обеспечивая надежность и обоснованность конечных результатов. Этот анализ закладывает прочную основу для последующей оптимизации и применения модели.

* 1. **Настройка эксперимента Анализ результатов**

После завершения обучения модель была оценена на тестовом наборе, который показал общую точность 85%. Такие результаты свидетельствуют о высокой надежности модели при распознавании нескольких эмоций. При анализе матрицы путаницы было обнаружено, что модель имеет высокую точность 90 % и 88 % для эмоций "счастье" и "грусть", соответственно. Однако для эмоций "удивление" и "отвращение" точность была значительно ниже - 75 % и 72 % соответственно. Это явление может быть связано со схожестью этих двух эмоций в выражениях лица и недостатком обучающих данных.

При дальнейшем анализе результатов запоминания и F1 обнаружилось, что существуют некоторые различия в производительности моделей при распознавании эмоций. Например, при распознавании эмоции "гнев" результат составляет 82 %, а при распознавании эмоции "страх" - только 76 %. Такая неравномерная производительность говорит о том, что в будущих исследованиях нам необходимо уделить особое внимание обучающим данным по этим эмоциям, чтобы повысить точность модели в этих категориях. Кроме того, отзывы пользователей также показывают, что модель демонстрирует хорошую производительность при распознавании эмоций в реальном времени, и пользователи в целом удовлетворены результатами распознавания эмоций.

В заключение, проведя всесторонний анализ результатов эксперимента, мы смогли подтвердить эффективность модели в задаче распознавания эмоций и определить конкретные направления для улучшения. Эти результаты служат важной основой для последующей оптимизации модели и расширения функций, обеспечивая более полное удовлетворение потребностей пользователей в практических приложениях.

* + 1. **Тестовая среда Точность и полнота**

Настройка тестовой среды имеет решающее значение для оценки моделей распознавания эмоций. В данном исследовании все эксперименты проводились в стандартной среде, чтобы обеспечить точность и воспроизводимость результатов. Для ускорения процесса обучения и вывода моделей использовалось высокопроизводительное компьютерное оборудование с мощными графическими процессорами. При этом были выбраны сценарии с различными условиями освещения и фоном, чтобы имитировать реальные приложения. Такое разнообразие тестовых сред позволяет оценить модели в различных условиях, чтобы убедиться в их адаптивности и стабильности.

Во время тестирования точность была ключевой метрикой для оценки работы модели. Сравнивая предсказания модели с реальными метками, рассчитывалась общая точность и точность для каждой категории эмоций. Результаты эксперимента показали, что модель достигла общей точности 85 %, с точностью 90 и 88 % для эмоций "счастье" и "грусть" соответственно. Однако для некоторых категорий эмоций, таких как "удивление" и "отвращение", точность оказалась несколько ниже - 75 и 72 % соответственно. Эти результаты говорят о том, что, хотя модель хорошо распознает большинство эмоций, ее еще можно улучшить.

Полнота - это согласованность работы модели при распознавании нескольких эмоций и условий окружающей среды. В экспериментах использовались различные наборы данных, такие как FER2013 и AffectNet, что позволяет модели сохранять хорошую производительность в различных контекстах. Отзывы пользователей и тесты в реальных сценариях применения показывают, что модель сохраняет высокий уровень точности при работе с выражениями лица со сложным фоном и окклюзиями, что свидетельствует о хорошей целостности модели.

Таким образом, проведя строгий эксперимент в стандартной тестовой среде, мы смогли всесторонне оценить точность и целостность модели распознавания эмоций. Эти результаты не только подтверждают эффективность модели, но и служат основой для дальнейшей оптимизации и усовершенствования, чтобы еще больше повысить ее эффективность в практических приложениях.

* + 1. **Тестовый набор данных Производительность модели при затенении**

В ходе экспериментов использовались изображения из наборов данных FER2013 и пользовательского, содержащие различную степень окклюзии. Процесс вывода модели показывает, что общая точность снижается в условиях окклюзии, особенно при распознавании эмоций "гнев" и "удивление". В частности, при полной окклюзии одного глаза или рта точность распознавания модели падала до 60 % для эмоции "удивление" и только до 65 % для эмоции "гнев". Эти результаты свидетельствуют о том, что частичная окклюзия черт лица существенно влияет на эффективность распознавания модели.

Для более глубокого анализа этого явления мы дополнительно изучили влияние различных уровней окклюзии на работу модели. В случае легкой окклюзии (например, палец слегка прикрывает рот) точность модели остается около 75 %, но в случае сильной окклюзии (например, шляпа полностью закрывает часть лица) точность падает до менее 50 %. Это говорит о том, что влияние окклюзии на распознавание эмоций тесно связано со степенью окклюзии.

В целом, несмотря на то, что модель хорошо работает в стандартных условиях, ухудшение производительности в ситуациях с окклюзией показывает ее ограниченность в реальных приложениях. Для решения этой проблемы в будущих исследованиях можно рассмотреть возможность введения в процесс обучения большего количества сценариев с окклюзией, чтобы улучшить способность модели адаптироваться к окклюзии. Кроме того, объединение мультимодальных данных (например, голоса и интонации) также может помочь повысить эффективность распознавания эмоций в сложных ситуациях.

# **Обсуждение**

Данное исследование дает представление о разработке и оценке моделей распознавания эмоций на мобильных устройствах, уделяя особое внимание производительности в условиях окклюзии. Результаты показывают, что, хотя модель демонстрирует высокую точность и устойчивость в широком диапазоне стандартных условий, ее производительность снижается при столкновении с ситуациями окклюзии, которые часто встречаются в реальном мире. Этот вывод подчеркивает необходимость учета влияния реальной среды на производительность модели в исследованиях по распознаванию эмоций.

Во-первых, общая точность модели в большинстве случаев является удовлетворительной, особенно при распознавании отдельных эмоций, таких как "счастье" и "грусть". Однако исследование также выявило значительные ограничения в распознавании некоторых эмоций, таких как "удивление" и "гнев". Это может быть связано с тонкими различиями в выражениях лиц для этих эмоций, особенно в случае окклюзии, когда эти детали могут быть скрыты окклюдером, что приводит к снижению способности модели к распознаванию. Поэтому в будущих исследованиях можно улучшить работу модели, увеличив количество обучающих образцов для этих эмоций, особенно в сложных сценах.

Во-вторых, влияние ситуаций окклюзии на распознавание эмоций подчеркивает важность набора данных. В данном исследовании, несмотря на то, что FER2013 и AffectNet предоставили богатые маркированные образцы, изображения с различными ситуациями окклюзии не были должным образом учтены во время обучения. Это говорит о том, что в будущем необходимо уделять больше внимания разнообразию образцов при сборе данных и обучении моделей, особенно в сложном контексте эмоционального выражения. Включение тренировочных данных с окклюзионными изображениями может помочь улучшить адаптацию модели в реалистичных сценариях.

Кроме того, данное исследование указывает на практичность и потенциальные трудности применения моделей распознавания эмоций в приложениях. В реальных сценариях выражение лица пользователя может быть закрыто по разным причинам, поэтому модель должна быть более надежной. По этой причине эффективным решением может стать объединение мультимодальных данных (например, голоса, интонации или языка тела). Такой комплексный подход может компенсировать отсутствие одного визуального сигнала и, таким образом, повысить точность распознавания эмоций.

Наконец, данное исследование дает представление о направлениях будущих исследований. При изучении методов распознавания эмоций исследователи должны обратить внимание на различия в выражении эмоций в разных культурных контекстах и на возможность применения в реальном времени на мобильных устройствах. В то же время разработка более персонализированной системы распознавания эмоций для пользователей разных возрастных групп, полов и этнических групп поможет улучшить пользовательский опыт и повысить универсальность системы. В целом, несмотря на то, что данное исследование позволило добиться определенного прогресса в области распознавания эмоций, остается еще много нерешенных вопросов и будущих направлений исследований, которые необходимо постоянно изучать и оптимизировать.

**ЗАКЛЮЧЕНИЕ**

При предварительной обработке изображений лица данные с шумом кренделя и гауссовским шумом не используются для операций фильтрации, а влияние различных фильтров на подавление шума анализируется путем вычисления пикового отношения сигнал/шум (PSNR), и делается вывод, что когда пиковое отношение сигнал/шум выше, эффект подавления шума лучше. В экспериментах по обнаружению лиц видно, что метод PCA чрезвычайно чувствителен к окклюзии, и точность распознавания по методу PCA быстро снижается при наличии окклюзии в солнечных очках. Предложенный одновременно метод LBP с преобразованием Габора под маской окклюзии повышает точность лишь на небольшую величину по сравнению с методом без маски, когда окклюзии нет. Однако при работе с замаскированными изображениями лица точность значительно повышается. Мы сравнили различные методы предварительной обработки и различные методы распознавания лиц при наличии шума в изображениях лиц. Различные эффекты при разных параметрах также сравниваются с методами распознавания эмоций. Обсуждается теория и применение механизма внимания, а также проводятся эксперименты с различными наборами данных для сравнения экспериментальных эффектов. Результаты экспериментов показывают, что алгоритм обнаружения лиц, предложенный в данной работе, может эффективно обнаруживать лица в условиях частичной окклюзии. В наборе данных AR частота обнаружения лиц без окклюзии, с окклюзией глаз и с окклюзией рта составляет 95,79%, 99,3% и 88,27% соответственно.

**СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ**

[1] Ekenel H K, Stiefelhagen R. Why id Facial Occlusion a Challenging Problem[C]. 3rd IARR/IEEE International Conference on Advances inBiometrics,2009:299-308.

[2] Oloyede M O, Hancke G P, Kapileswar N. Evaluating the effect of Occlusion in Face Recognition Systems[C]. IEEE AFRICON Conference – Science, Technology and Innovation for Africa,2017:1547-1551.

[3] Yan J, Lei Z, Wen L,et al.The Fastest Deformable Part Model for Object Detection[C]. 27th IEEE Conference on computer vision and pattern recognition (CVPR),2014:2497-2504.

[4] Deng J, Guo J, Ververas E, et al. RetinaFace: Single-shot Multi-level Face Localisation in the Wild[C]. IEEE/CVF Vision and Pattern Recognition (CVPR),2020:5202-5211.

[5] Chen Z, Wu X-J, Kittler J. A sparse regularized nuclear norm based matrix regression for face recognition with contiguous occlusion[J]. Pattern Recognition Leeter,2019, 125:494-299.

[6] Krizhevsky A, Sutskever I, Hinton G E, Image Net Classification with deep convolutional neural networks[J]. Communications of the Acm,2017,60(6):84-90.

[7] Wang M, Tian W, Kanf H,, et al. Biclustering gene expression data by an improved optimal algorithm[C]. International Conference on Mechatronics and Industrial Informatics (ICMII 2013),2013:2223-2226.