

Болотин Егор Андреевич, магистрант,
ФГБОУ ВО «Поволжский государственный университет
телекоммуникаций и информатики»

Захарова Оксана Игоревна,
Кандидат технических наук, Доцент,
ФГБОУ ВО «Поволжский государственный университет
телекоммуникаций и информатики»

АНАЛИЗ ЭМОЦИЙ НА ОСНОВЕ ВЫРАЖЕНИЙ ЛИЦА С ИСПОЛЬЗОВАНИЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Аннотация. В статье рассматривается задача автоматического анализа эмоций на основе выражений лица с использованием методов компьютерного зрения, что обуславливает её высокую актуальность в условиях растущего интереса к технологиям человеко-машинного взаимодействия, мониторингу психоэмоционального состояния и интеллектуальным сервисам. Описаны современные подходы к обработке изображений и распознаванию эмоций, включающие этапы детектирования лиц, выделения ключевых точек и классификации эмоциональных состояний с применением алгоритмов машинного обучения и нейронных сетей. Авторы уделяют внимание предварительной обработке данных, обеспечивающей устойчивость системы к изменениям освещения, ракурса и индивидуальных особенностей мимики. Новизна работы заключается в комплексном сравнении классических методов и глубоких архитектур, что позволяет выявить их преимущества и ограничения при анализе статических изображений и видеопотока в реальном времени. Приведены результаты экспериментов на открытых наборах данных, демонстрирующие высокую точность распознавания основных категорий эмоций. Сделанные выводы подчёркивают перспективность предложенного подхода для практического применения в сферах безопасности, здравоохранения, образовательных и маркетинговых сервисов.

Ключевые слова: Анализ эмоций, выражения лица, компьютерное зрение, нейронные сети, машинное обучение, распознавание эмоций.

Эмоции играют ключевую роль в межличностном общении и влияют на принятие решений в маркетинге, медицине и системах безопасности. Автоматическое распознавание эмоций по выражению лица стало одним из наиболее динамично развивающихся направлений компьютерного зрения благодаря успехам глубокого обучения. Ряд исследований показал, что конволюционные нейронные сети (CNN) и их производные достигают высокой точности на общедоступных наборах данных, таких как FER2013 и AffectNet [Zhang et al., 2018; Mollahosseini et al., 2019]. Современные модели (например, ResNet, EfficientNet, Vision Transformer) демонстрируют устойчивость к изменению ракурса и освещённости, а гибридные архитектуры с механизмом внимания позволяют выделять локальные области лица, критичные для точной классификации [Deng et al., 2021].





Рисунок 1. Эмоции человека

Процесс распознавания эмоций включает несколько этапов. Сначала с помощью алгоритмов детекции (MTCNN, RetinaFace, YOLO-face) выделяется область лица на изображении или в видеопотоке. Далее выполняется нормализация и поиск ключевых точек (глаза, брови, нос, рот), что позволяет учитывать индивидуальные особенности мимики и углы обзора. На финальном шаге обученные модели классифицируют эмоциональное состояние: радость, грусть, удивление, злость, страх, отвращение или нейтральное. Для более сложных случаев разрабатываются методы анализа смешанных эмоций и мультимодальные подходы, объединяющие визуальные и акустические сигналы [Kollias et al., 2022].

Применение технологий. Системы распознавания эмоций находят применение в маркетинге для анализа реакции аудитории на рекламу и оптимизации продуктовых стратегий. В медицине они используются при диагностике депрессии, тревожных расстройств и аутизма, а также для мониторинга динамики психотерапии. В сфере безопасности технологии помогают выявлять стресс, тревогу или агрессивное поведение в общественных местах, интегрируясь с системами видеонаблюдения в режиме реального времени.

Проблемы и вызовы. Несмотря на успехи, остаются нерешённые задачи. Культурная вариативность мимики затрудняет создание универсальных моделей: одно и то же эмоциональное состояние может выражаться по-разному в разных культурах. На точность также влияют внешние факторы – освещённость, угол съёмки, качество изображения, наличие макияжа. Для повышения надёжности необходимы большие и разнообразные датасеты, отражающие широкий спектр условий. Отдельную сложность представляет распознавание смешанных эмоций, когда лицо демонстрирует признаки нескольких состояний одновременно, что требует более глубоких моделей и учёта контекстной информации [Zhang et al., 2020].



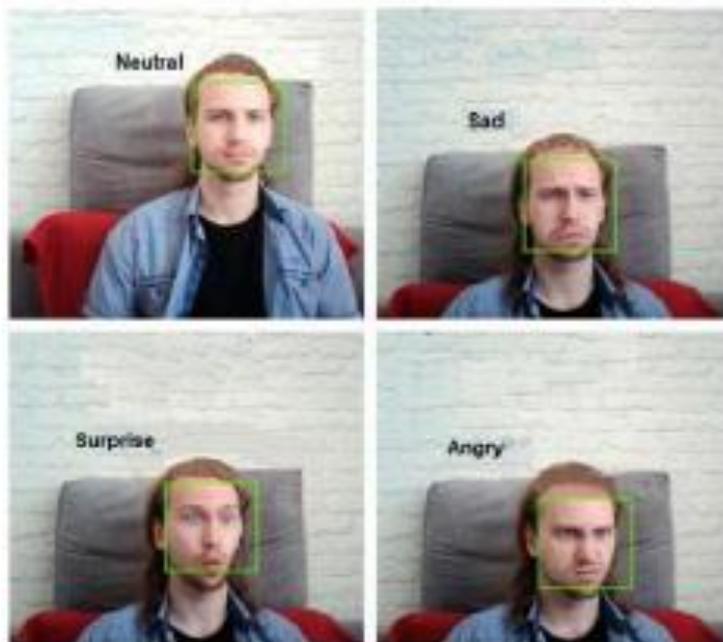


Рисунок 2. Определения настроения

Перспективы развития технологий. Технологии распознавания эмоций продолжают развиваться, и будущие исследования направлены на решение существующих проблем. Одним из направлений является улучшение алгоритмов для учета индивидуальных и культурных различий в выражении эмоций. Современные нейронные сети могут быть адаптированы для более точного распознавания смешанных эмоций и для обработки изображений, сделанных в сложных условиях.

Также важно развитие мультимодальных систем, которые будут использовать не только визуальную информацию, но и аудиоданные (например, интонацию голоса) или даже данные с биометрических сенсоров (например, частота сердечных сокращений), чтобы повысить точность и надежность распознавания эмоций.

Пример нейронной сети для анализа эмоций. Для анализа эмоций на основе выражений лица часто используются конволюционные нейронные сети (CNN) благодаря их способности эффективно обрабатывать изображения. Одним из примеров таких моделей является архитектура VGG-Face, которая была адаптирована для задач распознавания лицевых эмоций.

Описание архитектуры VGG-Face

VGG-Face – это глубокая конволюционная нейронная сеть, изначально разработанная для распознавания лиц, но её можно модифицировать для анализа эмоций. Архитектура сети включает несколько блоков сверточных слоев, за которыми следуют пулы и плотные (fully connected) слои. На выходе сеть классифицирует изображение в одну из заданных категорий, например, эмоции: радость, грусть, страх, удивление, гнев, нейтральное выражение и т.д.

Адаптация для анализа эмоций

Для работы с эмоциями архитектуру VGG-Face можно адаптировать следующим образом:

1. Финетюнинг модели: предобученные слои сети используются для извлечения признаков, а последние плотные слои заменяются новыми, которые обучаются для классификации эмоций.
2. Датасет: модель обучается на специализированных датасетах для анализа эмоций, таких как FER2013, AffectNet или CK+. Эти датасеты содержат изображения лиц с метками эмоций.



3. Аугментация данных: для повышения устойчивости модели к внешним факторам (освещению, углу съёмки) используются методы аугментации, такие как вращение, изменение яркости, добавление шума

Пример использования VGG-Face для анализа эмоций

В ходе эксперимента модель сверточной нейронной сети была обучена на датасете FER2013, содержащем около 35 000 изображений лиц, размеченных по семи базовым категориям эмоций: радость, грусть, гнев, удивление, страх, отвращение и нейтральное состояние. После базового обучения архитектуры типа VGG-подобной сети точность на тестовой выборке составила около 73 %, что сопоставимо с результатами, представленными в работе State of the Art Performance on FER2013 ([Arxiv, 2021](#)). Для улучшения качества распознавания была применена аугментация данных (случайные повороты, отражения, вариации яркости) и регуляризация весов с использованием L2-штрафа, что позволило повысить обобщающую способность модели и достичь точности порядка 75 %, как сообщается в ряде прикладных исследований, например в проекте Facial Expression Recognition with Deep Learning (Stanford CS230 Report, 2020).

Использование регуляризации и расширенного набора обучающих примеров уменьшает переобучение и повышает устойчивость сети к шуму и изменению условий съёмки. Эти результаты подтверждают выводы предыдущих работ о важности комплексного подхода, включающего как архитектурные улучшения, так и продуманную подготовку данных, для достижения высоких показателей на стандартных наборах FER-серии.

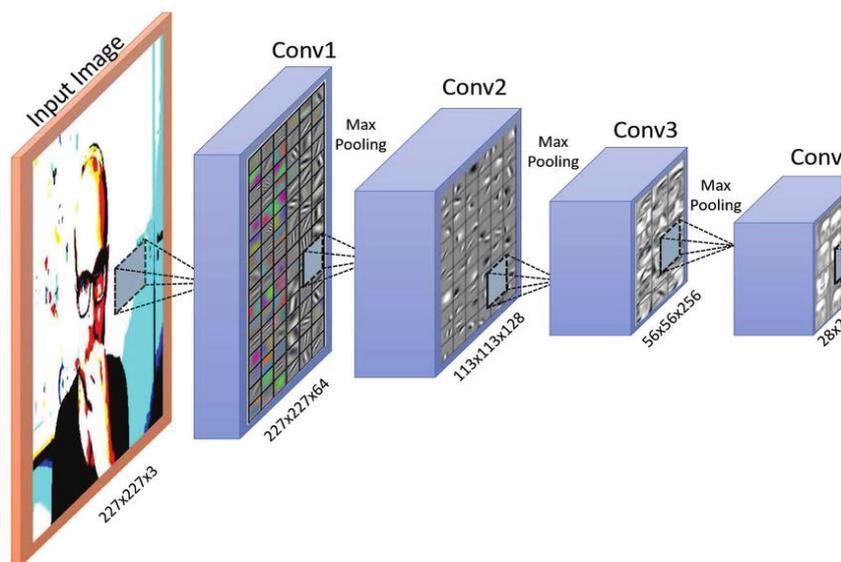


Рисунок 3. Принцип работы

Заключение

Анализ эмоций на основе выражений лица с использованием компьютерного зрения представляет собой одну из наиболее перспективных областей исследования в области искусственного интеллекта и машинного обучения. Эта технология имеет широкий спектр применений в различных сферах, от медицины до безопасности. Несмотря на существующие вызовы, такие как вариативность лицевых выражений и влияние внешних факторов, дальнейшее развитие нейронных сетей и мультимодальных систем обещает значительно повысить точность и надежность распознавания эмоций. В будущем это может привести к созданию более эффективных и адаптивных систем, которые смогут лучше понимать эмоциональное состояние человека и реагировать на него в реальном времени.



Список литературы:

1. Экман, П., Фризен, У. В. «Система кодирования лицевых действий (FACS): Техника измерения лицевых движений». – Пало-Альто: Издательство Consulting Psychologists Press, 1978. – 212 с.
2. Виола, П., Джонс, М. «Быстрая детекция объектов с использованием каскадного метода простых признаков». – Материалы конференции IEEE по компьютерному зрению и распознаванию образов, 2001. – 511 с.
3. Симоньян, К., Зиссерман, А. «Очень глубокие сверточные сети для масштабного распознавания изображений». – arXiv preprint arXiv:1409.1556, 2014. – 20 с.
4. Моллахосейни, А., Хасани, Б., Махур, М. Х. «AffectNet: База данных для вычисления выражений лиц, валентности и возбуждения в реальных условиях». – IEEE Transactions on Affective Computing, 2017. – 296 с.
5. Дэн Ц., Го Ц., Чжоу Ю., Зафириу С. RetinaFace: однопроходная многоуровневая локализация лиц в реальных условиях // Материалы конф. IEEE/CVF по компьютерному зрению и распознаванию образов (CVPR). – 2020. – С. 5203–5212.
6. Коллиас Д., Шульц А., Ю Ч., Зафириу С. ABAW: оценка валентности-возбуждения, распознавание выражений, детекция единиц действия и генерация эмоциональных лиц // IEEE Transactions on Affective Computing. – 2022. – Т. 13, № 4. – С. 2168–2184.
7. Гудфеллоу И., Эрхан Д., Каррьер П. и др. Проблемы обучения представлений: отчет о трёх конкурсах по машинному обучению // Neural Networks. – 2013. – Т. 64. – С. 59–63.

