

Классификация эмоций на портретных изображениях с использованием мультимодальной модели GPT-5.1

Кокухина Софья Александровна

Приамурский государственный университет им.Шолом-Алейхема

Студент

Аннотация

В статье рассматривается возможность автоматической классификации эмоций на портретных изображениях с использованием мультимодальной нейросети GPT-5.1. Модель анализирует входное изображение и определяет эмоциональное состояние человека, включая базовые категории: радость, грусть, удивление, злость, страх, нейтральное состояние. Оценивается корректность интерпретации выражений лица, устойчивость к изменениям ракурса и освещения, а также способность модели объяснять свой результат.

Ключевые слова: GPT-5.1, мультимодальные модели, классификация эмоций, обработка изображений, компьютерное зрение.

Classification of emotions in portrait images using the GPT-5.1 multimodal model

Kokukhina Sofya Alexandrovna

Sholom-Aleichem Priamursky State University

Student

Abstract

The article discusses the possibility of automatic classification of emotions in portrait images using a multimodal neural network GPT-5.1. The model analyzes the input image and determines the emotional state of a person, including the basic categories: joy, sadness, surprise, anger, fear, neutral state. The correctness of the interpretation of facial expressions, resistance to changes in angle and lighting, as well as the ability of the model to explain its result are evaluated.

Keywords: GPT-5.1, multimodal models, emotion classification, image processing, computer vision.

1. Введение

1.1 Актуальность

Распознавание эмоций на изображениях является одной из ключевых задач компьютерного зрения. Области применения включают психологические исследования, сервисы пользовательской обратной связи, системы контроля качества обучения, маркетинговые исследования, а также инструменты для безопасности и анализа поведения. Современные мультимодальные модели, такие как GPT-5.1, объединяют текстовый и

визуальный анализ, что позволяет использовать их как универсальный механизм интерпретации изображений без сложной отдельной модели CV.

1.2 Обзор исследований

В. Е. Гай рассмотрел задачу распознавания эмоций человека по изображению лица и описал процесс создания программного приложения, использующего алгоритмы глубокого обучения для автоматической классификации эмоциональных состояний, показав, что использование сверточных нейронных сетей позволяет добиться приемлемой точности в реальных условиях применения [1]. Е. В. Рюмина выполнила аналитический обзор современных методов распознавания эмоций по выражению лица и баз видеоданных, систематизировав подходы на основе классических признаков и глубоких нейронных сетей и указав на ключевые ограничения существующих решений, связанные с вариативностью мимики и условиями съёмки [2]. А. Б. Елизарова исследовала технологию распознавания эмоций человека по выражению лица с применением глубоких сверточных нейронных сетей, продемонстрировав, что переход к более глубоким архитектурам позволяет улучшить качество классификации по сравнению с традиционными методами компьютерного зрения [3]. И. О. Широкомяд провёл сравнение различных моделей машинного обучения для распознавания эмоций на изображениях, показав, что нейросетевые подходы превосходят по точности классические алгоритмы, однако требуют тщательного подбора архитектуры и параметров обучения [4]. А. П. Семёнова рассмотрела задачу определения ключевых точек лица для распознавания эмоций по мимическим микровыражениям, отметив, что точное выделение опорных точек существенно повышает устойчивость последующей классификации эмоциональных состояний, особенно в условиях шумов и небольших изменений позы [5].

1.3 Цель исследования

Цель исследования — оценить возможности нейросети GPT-5.1 в задаче автоматической классификации эмоций по портретным изображениям и определить факторы, влияющие на точность анализа.

2 Материалы и методы

В исследовании использовалась мультимодальная модель GPT-5.1, способная анализировать изображения и формировать текстовое описание эмоциональных состояний. Портреты вводились в модель в виде входных изображений, после чего выполнялась классификация эмоций и формирование объяснений результата.

Тестирование проводилось на трёх группах:

- стандартные портреты с чёткой мимикой;
- портреты с частичным закрытием лица или изменённым освещением;
- эмоционально сложные изображения (смешанная мимика, неоднозначные выражения).

Оценивалась точность определения категории эмоций, устойчивость к визуальным помехам и логичность интерпретации.

3. Результаты

3.1 Классификация эмоций на стандартных портретах

Первый сценарий включал анализ классического портретного изображения с выраженными эмоциями: радость, грусть, злость, удивление и нейтральное состояние. GPT-5.1 уверенно интерпретировала ключевые признаки выражения лица: поднятие бровей, улыбку, углы рта, напряжённость мышц глаз (Рис.1).



Рисунок 1 – Изображение для оценки эмоции

Промпт:

Analyze the emotion on this portrait. Provide the most likely category and explain your reasoning.

Модель давала точные и полные ответы, например: “Closed/relaxed eyes with the upper cheeks pushed up, which usually happens when someone is really happy or laughing.” При этом на нейтральных изображениях GPT-5.1 корректно определяла отсутствие выраженной эмоции (Рис. 2).

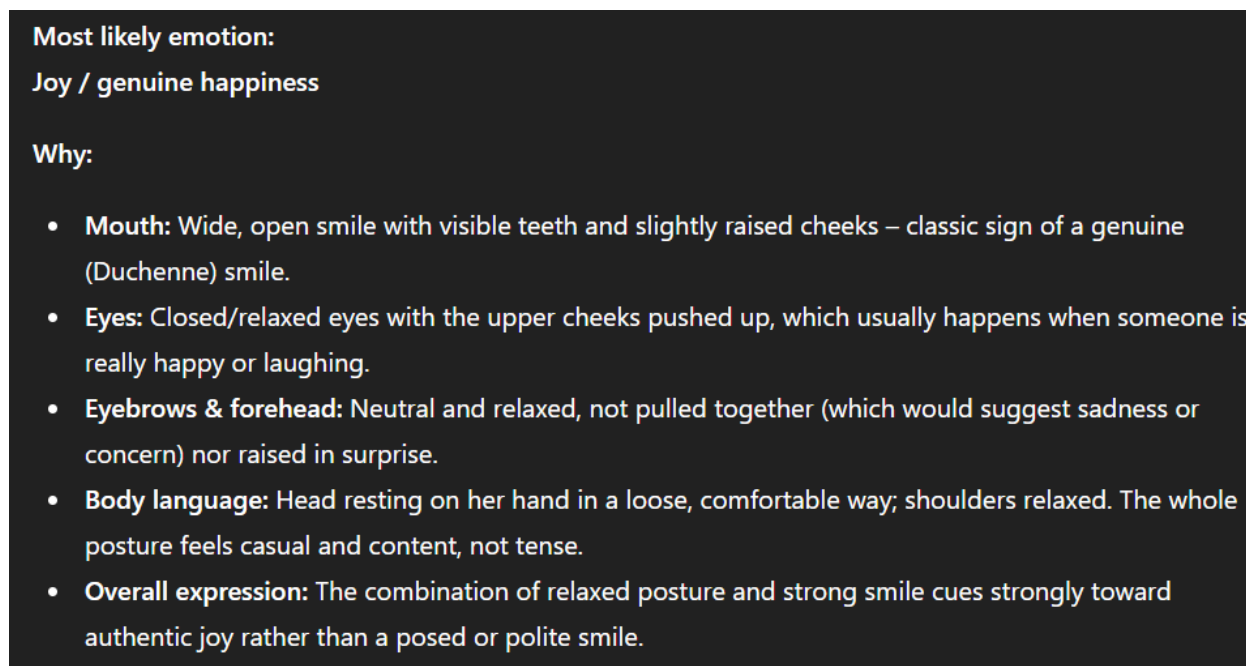


Рисунок 2 – Анализ эмоций на стандартном портрете

3.2 Анализ эмоций при частичных помехах

Во втором сценарии рассматривались фотографии, где лицо частично закрыто руками, одеждой, волосами или находится в условиях слабого освещения. Несмотря на искажённые данные, GPT-5.1 смогла выделить доступные признаки: оттенок кожи, положение глаз, остатки мимических линий (Рис.3).



Рисунок 3 – Пример изображения при помехах

Промпт:

Classify the emotional state on this portrait. The face may be partially occluded.

Модель выявила более интенсивную и потенциально агрессивную эмоциональную окраску. Ответ GPT-5.1 содержал акцент на подозрительности, напряжённости и скрытой агрессии, основываясь на сужении глаз, напряжении бровей и общей атмосфере изображения. Это указывает на то, что модель учитывает не только мимику, но и стилистические, контекстуальные элементы портрета, влияющие на настроение изображения. Это демонстрирует способность GPT-5.1 обрабатывать неполные данные (Рис. 4).

1. First Portrait (hand covering part of face)

Most likely emotion: Worry / anxiety / concern

Reasoning:

- The hand covering much of the face is a common self-soothing or stress gesture.
- The visible eye looks focused and tense, not relaxed.
- The lips appear slightly pressed together, which can indicate discomfort or concern.
- The overall posture suggests introspection or something weighing on the person's mind.

2. Second Portrait (person in hood, partially obscured face)

Most likely emotion: Intensity / suspicion / hostility

Reasoning:

- The person's eyes are narrowed and directed straight at the viewer, a typical sign of suspicion, anger, or strong focus.
- The face covering and hood add an impression of guardedness or intimidation.
- Brow muscles appear lowered and tightened, which commonly signals anger or determination.
- The dark environment reinforces a tense or confrontational emotional tone.

Рисунок 4 — Пример классификации эмоций при частичных помехах

3.3 Интерпретация сложных эмоциональных выражений

Третий сценарий включал изображения, где эмоции смешаны: удивление и радость, раздражение и усталость, печаль и страх. GPT-5.1 продемонстрировала способность распознавать комбинации признаков и формировать многоуровневый вывод (Рис.5).



Рисунок 5 – Фото для проверки модели со сложными эмоциями

Промпт:

Determine the emotional state. If multiple emotions are present, describe their proportion.

Модель продемонстрировала способность к распознаванию смешанных эмоциональных состояний, включающих оттенки удивления, восторга, ностальгии и лёгкой печали. Здесь GPT-5.1 не ограничилась одной категорией, а сформировала пропорциональный анализ нескольких эмоциональных компонентов, указав доминирующую эмоцию (awe/wonder) и её вклад в общее состояние. Такой тип анализа показывает, что модель способна проводить многослойную интерпретацию выражения лица, учитывая направление взгляда, степень раскрытости глаз, положение рта и жесты рук как элементы

сложного эмоционального паттерна. Такие интерпретации полезны в психологических приложениях и UX-исследованиях (Рис. 6).

Most likely emotional state:

A mix of awe, longing, and gentle sadness — with awe or wonder being the dominant emotion.

Proportions (approximate):

- Awe / wonder: ~60%

Her eyes are widened and slightly lifted upward, reflecting light in a way that suggests she is looking at something meaningful or beautiful. The open-mouth expression also often accompanies amazement or deep reflection.

- Longing / nostalgia: ~25%

The upward gaze combined with the hands holding her cheeks gives a sense of remembering or yearning for something distant—emotionally or in time.

- Sadness / melancholy: ~15%

There is a subtle heaviness around the mouth and cheeks, and the softness of her expression suggests traces of bittersweet emotion rather than pure sadness.

Reasoning:

- **Eyes:** Wide, moist, reflective, looking slightly upward — strong indicators of wonder or deep emotional reflection.
- **Mouth:** Slightly open, not forming a smile or frown, hinting at emotional overwhelm.
- **Hands on face:** A classic gesture of someone experiencing something emotionally intense, whether beautiful or poignant.
- **Overall expression:** Soft, vulnerable, emotionally rich — not one-dimensional.

Рисунок 6 – Сложные смешанные эмоции

3.4 Сравнение трёх типов сценариев

Сравнение трёх рассмотренных сценариев демонстрирует различия как в сложности интерпретации, так и в характере выводов, которые формирует модель GPT-5.1. В первом случае, где лицо было частично закрыто рукой, модель уверенно определила состояние тревоги и беспокойства, опираясь на такие признаки, как напряжённый взгляд, прижатые губы и жест самоприкосновения, который она интерпретировала как элемент саморегуляции. Даже при отсутствии значительной части лица модель устойчиво извлекла эмоциональные маркеры из ограниченной области, что показывает её способность работать с неполной информацией.

4 Выводы

Сопоставление результатов показывает, что GPT-5.1 наиболее уверенно работает с изображениями, где эмоции выражены комплексно и содержат богатый набор визуальных признаков. При этом в условиях помех или частичного скрывания лица модель сохраняет способность дать корректную интерпретацию, хотя её уверенность и детализация уменьшаются. Важно подчеркнуть, что GPT-5.1 демонстрирует не только классификационную

функцию, но и объяснительную, формируя обоснования, которые опираются на конкретные элементы мимики, позы и композиции изображения. Это делает модель применимой в задачах, где требуется интерпретируемый анализ эмоционального состояния.

Библиографический список

1. Гай В. Е., Зеленский В. П., Суворкин П. А. Распознавание эмоций человека по изображению лица // Информационные системы и технологии ИСТ-2021: сборник материалов XXVII Международной научно-технической конференции, Нижний Новгород, 23–24 апреля 2021. С. 831–837.
2. Рюмина Е. В., Карпов А. А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20, № 2. С. 163–176.
3. Елизарова А. Б. Технология распознавания эмоций человека по выражению лица с использованием глубоких сверточных нейронных сетей // Прикладная математика и информатика: современные исследования в области естественных и технических наук: Материалы VI Международной научно-практической конференции (школы-семинара) молодых ученых, Тольятти, 23–25 апреля 2020 года. С. 63–66.
4. Широкомяд И. О. Применение нейронных сетей для распознавания эмоций на фото по изображению лица // Наука и просвещение: актуальные вопросы, достижения и инновации: сборник статей X Международной научно-практической конференции. Пенза. 2023. С. 63–69.
5. Семенова А. П., Павлыш В. Н. Поиск ключевых точек лица для задачи распознавания эмоций // Информатика и кибернетика. 2021. № 1–2 (23–24). С. 59–64.