

## Классификация микровыражений на основе оптического потока с учётом гендерных особенностей

*Н.А. Мишин*

*Южный Федеральный Университет, Ростов-на-Дону*

**Аннотация:** В данной работе представляется методология распознавания микровыражений на основе оптического потока и архитектуры YOLOv11. В отличие от ранее предложенного подхода, ориентированного на детекцию: «есть микровыражение/нет микровыражения», в настоящем исследовании производится полноценная классификация различных типов микровыражений. Кроме того, учитывается половой признак испытуемых, поскольку вариации мимических проявлений могут различаться у мужчин и женщин. Для вычисления оптического потока применяется новый алгоритм, а в качестве инструмента классификации – модель YOLOv11, адаптированная под задачу анализа выражений лица. С целью нивелирования сложностей, связанных с дисбалансом классов, при оценке качества экспериментов используется Micro ROC-AUC-метрика. Результаты экспериментов показывают, что предложенный подход достигает конкурентоспособных показателей точности классификации даже при наличии существенного дисбаланса в выборке.

**Ключевые слова:** микровыражения, распознавание образов, оптический поток, YOLOv11.

### Введение

Эмоции и их внешние проявления играют важную роль в межличностном общении, способствуя пониманию намерений и чувств людей. Микровыражения, являющиеся кратковременными и зачастую неосознаваемыми изменениями мимики лица, давно привлекают внимание исследователей в различных областях – от психологии до систем безопасности [1].

В предыдущих исследованиях была предложена методика двоичной детекции микровыражений на основе оптического потока и разбиения лица на регионы, учитывающего контуры лицевых мышц [2]. Результаты показали, что точная сегментация лица способствует повышению эффективности обнаружения микровыражений. В рамках данного исследования мы

переходим к задаче многоклассового распознавания (6 классов), используя новую методику вычисления оптического потока, представленную в статье, находящейся в печати. При распознавании эмоций учитывается половой признак испытуемых (мужчины/женщины). Мы исходим из гипотезы, что гендерные различия (а в перспективе и этнические особенности) могут влиять на характер мимических паттернов, и, следовательно, на точность их распознавания.

Для классификации микровыражений выбрана архитектура YOLOv11, модифицированная под задачу анализа и классификации выражений лица. Несмотря на сложность задачи и наличие значительного дисбаланса классов в доступных наборах данных, нам удалось получить результаты, которые могут служить основой для дальнейших улучшений, особенно с учётом более диверсифицированных данных.

## **1. Используемые данные и предварительная обработка**

### **1.1. Датасет**

В основе исследования лежит комбинация видеозаписей, из набора CAS(ME)3 [3]. В совокупности в датасете представлены 6 основных классов микровыражений (счастье, грусть, удивление, гнев, отвращение, страх), при этом распределение классов неравномерно (рис. 1).

### **1.2. Гендерное разделение и этническая гетерогенность**

Для проверки гипотезы о том, что пол может влиять на характер и проявление микровыражений, была собрана информация о поле участников. Поскольку ручная разметка в масштабном датасете усложнена, распознавание пола производилось автоматически с помощью библиотеки DeepFace [4], применяемой к первым кадрам видеороликов. Модель DeepFace использует нейронные сети, предобученные на крупных базах лицевых изображений, что

---

обеспечивает высокую точность (более 99%) в то время, как человек в среднем показывает точность 97.5%.

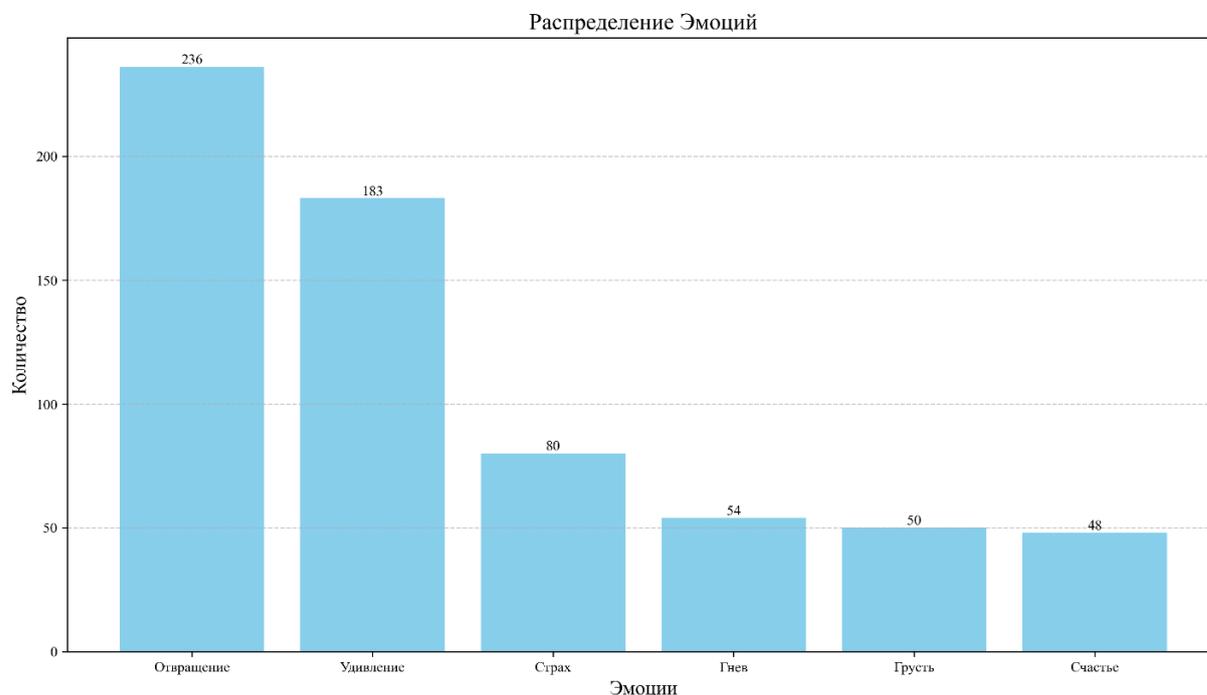


Рис. 1. – Распределение эмоций

В случаях, когда показатель уверенности (confidence) в определении пола оказывался ниже заданного порога, мы исключали соответствующие ролики из дальнейшего анализа. На основе итоговой классификации по полу мы сформировали две подвыборки: ролики с мужскими лицами (40 человек) и женскими лицами (54 человека).

### 1.3. Предобработка видео

**Перевод в чёрно-белый (монохромный) формат.** Все кадры изначально преобразуются из цветного RGB в градации серого, что снижает влияние цветовых вариаций и упрощает последующую обработку.

**Нормализация яркости и контраста.** Для минимизации влияния неравномерного освещения и контраста к кадрам применяется выравнивание гистограмм [5]. Это помогает выделить мелкие детали лицевых движений.

**Определение лицевых регионов.** Используется библиотека MediaPipe [6], которая на каждом кадре выделяет 478 ключевых точек лица. На основании этих точек формируется разбиение лица на 42 региона, соответствующие основным группам лицевых мышц. Каждый регион рассматривается как отдельная область, внутри которой будет усредняться величина оптического потока.

**Вычисление оптического потока.** Для каждой последовательной пары кадров из видеороликов, рассчитывается оптический поток с помощью нового алгоритма, который будет опубликован в ближайшее время [7]. Для каждого региона  $R_i$ , состоящего из множества пикселей, вычисляются усредненные значения модулей и углов оптического потока:

$$\bar{r}_i = \frac{1}{|R_i|} \sum_{j \in R_i} r_j, \quad \bar{\theta}_i = \frac{1}{|R_i|} \sum_{j \in R_i} \theta_j,$$

где  $r_i$  – это модуль вектора оптического потока, а  $\theta_i$  – угол направления вектора. Однако, в дальнейшем мы сосредотачиваемся только на угловой компоненте  $\theta$ , поскольку она показала большую устойчивость к вариациям интенсивности движения при анализе микровыражений.

**Дискретизация угловой компоненты.** Полученные значения  $\theta$  покрывают диапазон от 0 до  $2\pi$ . Для упрощения дальнейшего анализа и снижения чувствительности к незначительным колебаниям угла, этот интервал разбивается на 8 равных секторов. Формально, если  $\theta_i$  – усреднённый угол в  $i$ -м регионе, то после дискретизации:

$$\theta'_i = \frac{2\pi}{8} \left\lfloor \frac{8\theta_i}{2\pi} \right\rfloor.$$

Таким образом, в каждом регионе угол округляется до одного из восьми значений  $\frac{2\pi}{8}j$ , где  $j \in \{0, 1, \dots, 7\}$ , что снижает чувствительность модели к малым флуктуациям оптического потока [8]. Далее, дискретизированные значения масштабируются до диапазона  $[0, 255]$  для приведения в формат, подходящий для нейросети.

**Подготовка входных данных для YOLO.** В результате описанных шагов каждый кадр представляется вектором из 42 компонентов  $\{\theta'_1, \theta'_2, \dots, \theta'_{42}\}$ . Для учёта динамики во времени мы рассматриваем окно из 64 последовательных кадров, объединяя соответствующие векторы по кадрам в двумерный массив  $42 \times 64$ .

В таком массиве:

- индекс строки (от 1 до 42) соответствует конкретному лицевому региону;
- индекс столбца (от 1 до 64) соответствует номеру кадра во временной последовательности.

Подобное представление – это фактически «изображение» размерности  $42 \times 64$  пикселя, где значение яркости пикселя соответствует сектору угла оптического потока в данном регионе и в данный момент времени (рис. 2).

Таким образом, к моменту подачи данных в сеть YOLO каждая временная выборка (окно из 64 кадров) трансформируется в «псевдоизображение», на котором архитектура свёрточной нейронной сети способна эффективно учиться распознавать пространственно-временные паттерны микровыражений.

## 2. Архитектура YOLOv11-S и стратегия обучения

### 2.1. Общие сведения об архитектуре YOLO

YOLO (You Only Look Once) изначально была предложена как одностадийный метод детекции объектов, сочетающий в себе высокую скорость обработки и конкурентоспособное качество распознавания [9]. Начиная с ранних версий, разработка эволюционировала за счёт добавления остаточных блоков (Res-blocks), улучшения механизма skip-соединений, интеграции новых блоков нормализации и увеличения глубины сети для более точного прогнозирования.

---

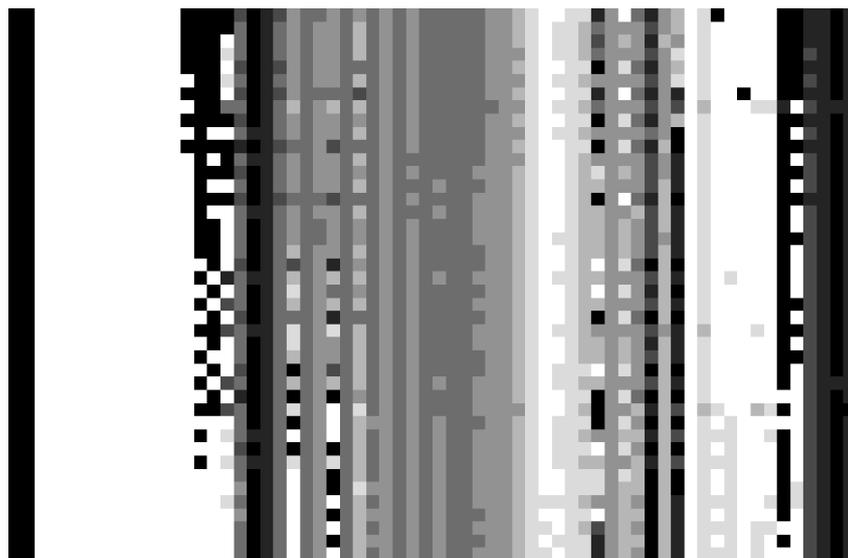


Рис. 2. – Пример входных данных

В данной работе применена облегчённая версия модели, условно обозначаемая как YOLOv11-S (Small), которая характеризуется:

- уменьшенным числом фильтров и слоёв по сравнению со старшими модификациями (M, L, XL);
- сокращённым объёмом параметров (около 9 млн.), что упрощает обучение и снижает требования к вычислительным ресурсам;
- возможностью достигать приемлемого качества даже при относительно небольшом количестве данных благодаря тщательному подбору архитектурных гиперпараметров.

Хотя YOLO в классическом виде ориентирована на детекцию, в нашей задаче сеть переориентируется на многоклассовую классификацию микровыражений.

### **3. Эксперименты и результаты**

#### **3.1. Метрики и учёт дисбаланса**

При оценке качества классификации микровыражений в условиях значительного дисбаланса классов использование стандартных метрик, таких

---

как точность (Accuracy) или F1-мера, может давать искажённые результаты. В связи с этим в данной работе была выбрана метрика ROC-AUC [10], а именно её вариант с микроусреднением (Micro-average ROC-AUC).

Пусть у нас есть многоклассовая задача классификации с классами, и каждый образец представлен в формате one-vs-all [11]. Рассматривается датасет, содержащий  $N$  примеров. Для каждого примера  $i$  имеется:

Вектор истинных меток  $y_i \in \{0,1\}^c$ , где  $y_{ic} = 1$ , если пример принадлежит классу  $c$ , и 0 в противном случае.

Вектор предсказанных вероятностей  $\hat{y}_i \in [0,1]^c$ , где  $\hat{y}_{ic}$  – предсказанная вероятность принадлежности к классу  $c$ ;

Для расчета Micro Average ROC-AUC:

- объединяем предсказания всех классов: формируем единый список пар (предсказанная вероятность, истинная метка) для всех классов;
- строим общий ROC-кривую: сортируем эти пары по убыванию вероятности и последовательно вычисляем точки ROC-кривой;
- вычисляем AUC по полученной кривой.

ROC-кривая строится путем изменения порога классификации и вычисления значений True Positive Rate (TPR) и False Positive Rate (FPR):

$$TPR = \frac{TP}{TP+FN}, FPR = \frac{FP}{FP+TN},$$

где:

- $TP$  (True Positives) – количество правильно предсказанных положительных классов;
  - $FP$  (False Positives) – количество ложно предсказанных положительных классов;
  - $FN$  (False Negatives) – количество пропущенных положительных классов;
-

- $TN$  (True Negatives) – количество правильно предсказанных отрицательных классов.

Мы вычисляем эти показатели, последовательно изменяя порог классификации от 1 до 0.

После построения ROC-кривой вычисляется площадь под кривой (AUC) численным интегрированием, например, методом трапеций:

$$AUC = \sum_{i=1}^{n-1} (FPR_{i+1} - FPR_i) \cdot \frac{TPR_i + TPR_{i+1}}{2},$$

где  $n$  – количество точек на ROC-кривой.

Метрика ROC-AUC используется для оценки качества бинарных и многоклассовых классификаторов, измеряя способность модели различать классы. Чем больше площадь под ROC-кривой (AUC), тем лучше модель справляется с классификацией.

В многоклассовом случае возможны два варианта усреднения:

- Macro ROC-AUC – вычисляет AUC отдельно для каждого класса и затем усредняет их показатели (не учитывает размер классов);
- Micro ROC-AUC – агрегирует все предсказания воедино перед построением ROC-кривой (учитывает размер классов).

Метрики Micro Average ROC-AUC и Macro Average ROC-AUC имеют принципиальные различия в способе вычисления и интерпретации.

### **Метод усреднения**

Micro Average ROC-AUC рассчитывает показатели True Positive Rate (TPR) и False Positive Rate (FPR) глобально, объединяя предсказания всех классов в единый набор данных. Таким образом, он строит единую ROC-кривую для всей выборки.

Macro Average ROC-AUC, напротив, сначала вычисляет AUC для каждого класса отдельно, а затем берет их среднее арифметическое. Это

позволяет учитывать вклад каждого класса независимо от частоты его встречаемости.

### **Учет дисбаланса классов**

Micro Average ROC-AUC учитывает дисбаланс классов, так как его расчет основан на всех предсказанных примерах. В этом случае классы с большим числом объектов оказывают большее влияние на итоговый результат.

Macro Average ROC-AUC, напротив, не учитывает дисбаланс, поскольку при усреднении каждому классу придается равный вес, вне зависимости от количества примеров, относящихся к нему.

### **Интерпретация**

Micro Average ROC-AUC отражает общую точность модели по всему датасету, без разделения на отдельные классы. Он дает представление о том, насколько модель хорошо справляется со всей выборкой в целом.

Macro Average ROC-AUC показывает, насколько хорошо модель предсказывает каждый класс в отдельности, усредняя результаты. Этот показатель более чувствителен к ошибкам на редких классах.

## **3.2. Результаты**

Результаты экспериментов на различных данных и параметрах представлены в таблице 1. Из таблицы 1 видно, что применение нового метода вычисления оптического потока в сочетании с архитектурой YOLOv11 позволяет добиться увеличения целевого показателя по сравнению ранее использованными подходами. Несмотря на наличие дисбаланса классов и вариативность в проявлениях микровыражений между мужчинами и женщинами, модель демонстрирует стабильно высокую способность к обобщению.

Таблица №1

Результаты экспериментов (большее значение метрики – лучше)

---

---

Тип данных	Значение ROC_AUC
Муж.+жен. без дискретизации угла	0,57
Муж.+жен. с дискретизацией угла	0,59
Муж. без дискретизации угла	0,53
Муж. с дискретизацией угла	0,6
Жен. без дискретизации угла	0,55
Жен. с дискретизацией угла	0,62

Дополнительно было проведено исследование, показывающее, что учёт гендерного признака (мужчина/женщина) с последующим обучением отдельных моделей, также увеличило целевую метрику несмотря на то, что количество обучающих данных сократилось почти вдвое.

#### 4. Выводы и дальнейшие исследования

В данной статье предложена методика распознавания нескольких классов микровыражений, использующая усовершенствованный алгоритм вычисления оптического потока и нейросетевую архитектуру YOLOv11. Одним из ключевых аспектов исследования является учёт гендерных различий. Было продемонстрировано, что разделение данных по полу и обучение отдельных моделей для мужчин и женщин улучшает качество классификации. Несмотря на снижение общего количества данных при разбиении на подгруппы, модели показали более высокие значения ROC-AUC, что свидетельствует о значимых для распознавания различиях в мимических образах у мужчин и женщин.

Дополнительно была проведена оценка эффективности различных подходов к обработке угловой компоненты оптического потока. Результаты показали, что дискретизация углового сигнала повышает стабильность модели, особенно в условиях значительных вариаций выражений лиц.

Используемая архитектура YOLOv11 продемонстрировала высокую производительность в задаче классификации микровыражений. Благодаря адаптации архитектуры под анализ временных последовательностей, модель смогла эффективно обрабатывать пространственно-временные паттерны мимики, несмотря на дисбаланс классов в датасете. Перспективными направлениями будущих исследований являются:

- изучение влияния этнических особенностей на выражение эмоций;
- разработка улучшенных стратегий обучения, включающих методы балансировки классов, такие как генеративные модели;
- внедрение временных зависимостей в модель за счёт интеграции рекуррентных сетей или трансформеров.

Представленные результаты подтверждают эффективность предложенного подхода и открывают возможности для дальнейшего совершенствования систем автоматического распознавания эмоций.

### Литература

1. Yan W. J. et al. How fast are the leaked facial expressions: The duration of micro-expressions // Journal of Nonverbal Behavior, 2013. V. 37. pp. 217-230
2. Мишин Н. А. Детекция микровыражений в видеоряде с помощью оптического потока // Информационные технологии и математическое моделирование (ИТММ-2024): труды Международной научно-практической конференции пос. Дивноморское, 25 августа – 30 августа 2024 г. – Ростов-на-Дону, 2024. С. 138–143.
3. Li J. et al. CAS (ME) 3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. V. 45. №. 3. pp. 2782-2800.

4. Serengil S. I., Ozpinar A. Hyperextended lightface: A facial attribute analysis framework // 2021 International Conference on Engineering and Emerging Technologies (ICEET). – IEEE, 2021. pp. 1-4.
5. Jha K. et al. A review on image enhancement techniques using histogram equalization //AIDE-2023 and PCES-2023, 2023. V. 497. pp. 497-502.
6. Lugaresi C. et al. Mediapipe: A framework for building perception pipelines // arXiv preprint. ArXiv: 1906.08172, 2019. URL: [arxiv.org/abs/1906.08172](https://arxiv.org/abs/1906.08172).
7. Бутакова, М. А., Щербань, И. В., Мишин, Н. А., Белявский Г. И. Два метода оценки оптического потока по видеоряду изображений // Искусственный интеллект и принятие решений. – 2025. В публикации.
8. Liu Y. J. et al. A main directional mean optical flow feature for spontaneous micro-expression recognition // IEEE Transactions on Affective Computing. – 2015. – V. 7. – №. 4. – pp. 299-310.
9. Khanam R., Hussain M. Yolov11: An overview of the key architectural enhancements // arXiv preprint. arXiv: 2410.17725, 2024. URL: [arxiv.org/abs/2410.17725](https://arxiv.org/abs/2410.17725).
10. Hand D. J., Till R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems //Machine learning, 2001. V. 45. pp. 171-186.
11. Bishop C. M., Nasrabadi N. M. Pattern recognition and machine learning. – New York: springer, 2006. V. 4. №. 4. P. 738.

### References

1. Yan W. J. et al. How fast are the leaked facial expressions: The duration of micro-expressions Journal of Nonverbal Behavior, 2013. V. 37. pp. 217-230.
  2. Mishin N. A. Informacionnye tekhnologii i matematicheskoe modelirovanie (ITMM-2024): trudy Mezhdunarodnoj nauchno-prakticheskoy
-

konferencii pos. Divnomorskoe, 25 avgusta- 30 avgusta 2024 g. Rostov-on-Don, 2024. pp. 138–143.

3. Li J. et al. CAS (ME) 3: A third generation facial spontaneous micro-expression database with depth information and high ecological validity IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. V. 45. №. 3. pp. 2782-2800.

4. Serengil S. I., Ozpinar A. Hyperextended lightface: A facial attribute analysis framework 2021 International Conference on Engineering and Emerging Technologies (ICEET). IEEE, 2021. pp. 1-4.

5. Jha K. et al. A review on image enhancement techniques using histogram equalization AIDE-2023 and PCES-2023, 2023. V. 497. pp. 497-502.

6. Lugaresi C. et al. Mediapipe: A framework for building perception pipelines. arXiv preprint. arXiv:1906.08172, 2019. URL: [arxiv.org/abs/1906.08172](https://arxiv.org/abs/1906.08172).

7. Butakova, M. A., Shcherban', I. V., Mishin, N. A., Belyavskij G. I. Iskusstvennyj intellekt i prinyatie reshenij. 2025. V publikacii.

8. Liu Y. J. et al. A main directional mean optical flow feature for spontaneous micro-expression recognition. IEEE Transactions on Affective Computing. 2015. V. 7. №. 4. pp. 299-310.

9. Khanam R., Hussain M. Yolov11: An overview of the key architectural enhancements arXiv preprint. arXiv:2410.17725, 2024. URL: [arxiv.org/abs/2410.17725](https://arxiv.org/abs/2410.17725).

10. Hand D. J., Till R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems Machine learning, 2001. V. 45. pp. 171-186.

11. Bishop C. M., Nasrabadi N. M. Pattern recognition and machine learning. New York: springer, 2006. V. 4. №. 4. P. 738.

**Дата поступления: 1.02.2025**

**Дата публикации: 3.03.2025**

---