

Minion Lab: это не мем это моя жизнь

Гончаренко Полина, Казумьян Максим, Загайнов Артём,
Степанов Михаил, Кузнецов Степан

1 Введение

В проекте решается задача сопоставления эмоции человека с изображением миньона. По видеопотоку с лицом человека необходимо определить эмоциональное состояние и отнести его к одному из восьми заранее заданных классов.

Проект вдохновлён интернет-мемами и примерами из социальных сетей, но также имеет потенциал для использования в интерактивных приложениях и развлекательных сервисах.

2 Данные

Для каждого класса было записано по 5 коротких видео (3–10 секунд). Видео разбиваются на кадры с шагом 1 кадр, и каждый кадр используется как отдельный обучающий пример.

Таким образом, из небольшого количества видео формируется достаточно крупный датасет изображений.

3 Метод и пайплайн

Мы построили собственный пайплайн обработки данных, включающий несколько этапов:

3.1 Этапы обработки

1. **Разбиение видео на кадры.**
2. **Аугментация** (повороты, небольшие искажения, изменение яркости).
3. **Извлечение признаков с помощью MediaPipe:** координаты ключевых точек лица и позы.
4. **Перемешивание и разделение** на обучающую и тестовую выборки.
5. **Масштабирование признаков** (fit только на train).
6. **РСА** для уменьшения размерности и ускорения обучения.
7. **Обучение SVM-классификатора.**

3.2 Инференс

При работе модели в реальном времени:

- кадр подаётся в MediaPipe;
- признаки масштабируются и сжимаются PCA;
- SVM выдаёт класс миньона;
- результат отображается пользователю.

4 Результаты

4.1 Матрица ошибок

	1	2	3	4	5	6	7	8
1	140	1	1	2	0	2	0	0
2	0	85	1	4	0	0	0	0
3	1	0	121	1	4	2	0	0
4	0	0	0	66	0	0	0	0
5	0	0	6	2	105	1	3	0
6	8	0	1	1	0	127	0	0
7	1	0	3	1	0	1	128	0
8	1	0	0	8	0	2	0	28

4.2 Классификационный отчёт

Класс	Precision	Recall	F1	Support
1	0.93	0.96	0.94	146
2	0.99	0.94	0.97	90
3	0.91	0.94	0.92	129
4	0.78	1.00	0.87	66
5	0.96	0.90	0.93	117
6	0.94	0.93	0.93	137
7	0.98	0.96	0.97	134
8	1.00	0.72	0.84	39
Accuracy	0.93			858

5 Анализ результатов

Основные наблюдения:

- Общая точность модели составляет **93%**, что является хорошим результатом.
- Большинство классов распознаются уверенно (F1 выше 0.9).

- **Класс 4** имеет высокий recall, но более низкий precision — модель иногда относит к нему лишние примеры.
- **Класс 8** имеет идеальный precision, но низкий recall — часть примеров класса пропускается.
- Эти проблемы, скорее всего, связаны с недостатком данных для классов 4 и 8.

6 Перспективы

- Добавить больше видео для слабых классов.
- Расширить набор миньонов и эмоций.
- Использовать распознавание нескольких лиц для «групповых миньонов».
- Реализовать добавление новых мемов пользователем.

7 Заключение

Мы разработали рабочую систему, которая по лицу человека определяет эмоцию и сопоставляет её с миньоном. Пайплайн на основе MediaPipe, PCA и SVM показал высокое качество и устойчивость. Проект может быть расширен и улучшен за счёт увеличения датасета и доработки отдельных этапов.