

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ РАСПОЗНОВАНИЯ ЭМОЦИЙ

А. Н. Гаврилин¹⁾, А. А. Лобанков²⁾ (научный руководитель)

¹⁾ магистр, Самарский университет, Самара, Российская Федерация, gavrilka100@mail.ru

²⁾ доцент, Самарский университет, Самара, Российская Федерация, mart1989@mail.ru

Поскольку количество аудио и видео информации в мире с каждым годом растёт, поэтому требуется новые алгоритмы её обработки. По этой причине проблема распознавания эмоций в наш век остается достаточно актуальной. Существует три основных метода классификации эмоций: классическая оценка лица с изображения, оценка изображения с применением глубокого обучения, или оценка речи. Ведь распознавание эмоций студентов на лекции поможет понять их вовлеченность в учебный процесс и соответственно изменить подход обучения в лучшую сторону или оценка состояния человека по голосу может спасательным службам успокоить его. В статье приведен сравнительный анализ основных алгоритмов распознавания эмоций, выделены достоинства и недостатки каждого алгоритма.

Ключевые слова: ЭМОЦИИ; распознавание; ключевые точки; глубокое обучение.

COMPARATIVE ANALYSIS OF EMOTION RECOGNITION METHODS

A. N. Gavrilin¹⁾, A. A. Lobankov²⁾ (supervisor)

¹⁾ Master, Samara University, Samara, Russian Federation, gavrilka100@mail.ru

²⁾ Associate Professor, Samara University, Samara, Russian Federation, mart1989@mail.ru

Since the amount of audio and video information in the world is growing every year, therefore, new algorithms for its processing are required. For this reason, the problem of recognizing emotions in our century remains quite relevant. There are three main methods for classifying emotions: classical face-from-image evaluation, deep-learning image evaluation, or speech evaluation. After all, recognizing the emotions of students at a lecture will help to understand their involvement in the educational process and accordingly change the approach of teaching for the better, or the assessment of a person's condition by voice can help rescue services reassure him. The article provides a comparative analysis of the main algorithms for recognizing emotions, highlighting the advantages and disadvantages of each algorithm.

Keywords: Emotions; recognition; key points; deep learning.

Стандартным способом оценки эмоционального спектра на изображении, является алгоритм ключевых точек (facial landmarks). На изображении строится система координат с помощью одного из алгоритмов: PDM (Point Distribution Model), CLM (Constrained Local Model), ААМ (Active Appearance Model). Такая привязка позволяет захватить мимику объекта. Далее построенные точки нормализуются и передаются непосредственно в классификатор. В роли классификатора может выступать метод опорных векторов (SVM) или, например, Random Forest.

К примеру, алгоритм АММ использует сервис FaceReader нидерландской компании Noldus Information Technology. Программа способна интерпретировать микроэкспрессии лица, распределяя их по семи основным категориям: радость, грусть, гнев, удивление, страх, отвращение и нейтральная (neutral). Кроме того, FaceReader умеет с достаточно высокой точностью определять по лицам возраст и пол человека.

Компания использует метод Active Template, заключающейся в наложении на изображение лица деформируемого шаблона, и методе Active Appearance Model (АММ),

позволяющем создавать искусственную модель лица по контрольным точкам с учетом деталей поверхности. Точность считывания эмоции составляет порядка 89 % [1]. Достоинствами данного метода являются простота по сравнению с иными методами, высокая точность определения эмоций. К недостаткам относится сильная зависимость точности от положения лица, освещения и качества изображения, что приводит к увеличению количества расчётов и уменьшению точности.

На рисунке 1 представлен пример использования алгоритма ААМ.

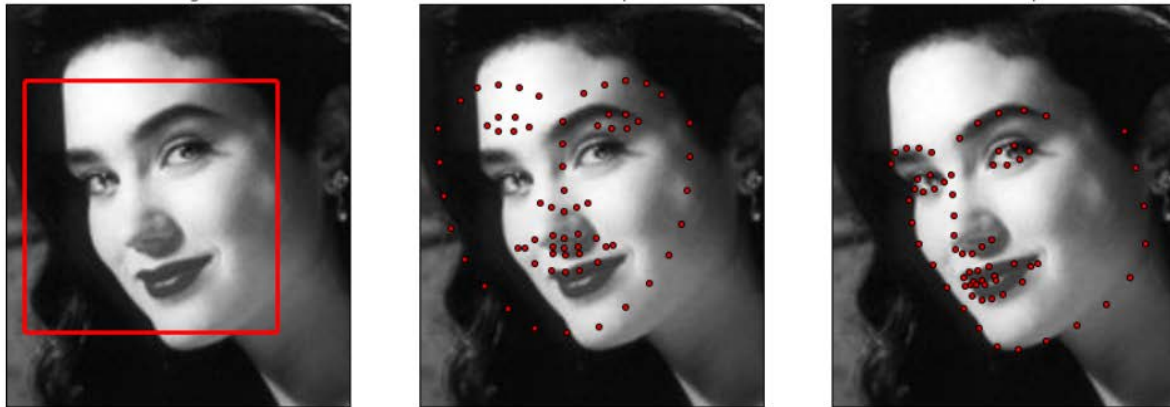


Рисунок 1 – Пример использования алгоритма ААМ

Иным способом классификации эмоций по изображению может быть применение CNN с методом глубокого обучения. Обычно сеть представлена в виде нескольких слоев: свертки, подвыборки и полносвязного слоя. Применение такой сети несет необходимость дополнительной подготовки данных для обучения, а также возможность переобучить используемую сеть что влечёт к снижению точности оценивания реальных данных. Пример CNN сети представлен на рисунке 2, точность такой сети в реальных условиях равно 60–90 % [2]. К достоинствам данного метода можно отнести меньшую зависимость от освещения и положения лица на изображении, также снижается требование к качеству, поскольку изображение подвергается дополнительной обработке. К недостаткам относится возможность некорректного обучения сети, что приведет к значительному снижению результата.

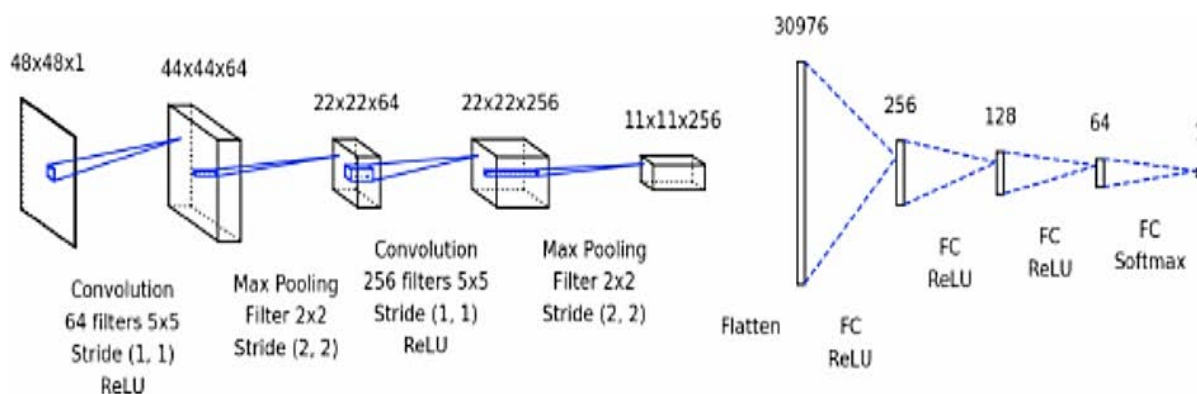


Рисунок 2 – Пример CNN сети с методом глубокого обучения

С помощью аудио данных эмоции классифицируются следующим образом: Полученные данные представляются в виде спектрограммы, затем подвергаются преобразованию Фурье и конвертируются в изображение. Затем полученные

изображения проходят через рекуррентную сеть. Такой метод обработки речевых сигналов приведен в статье «A breakthrough in Speech emotion recognition using Deep Retinal Convolution Neural Networks», точность такой сети составила 90–99 % [3, 4]. Достоинства и недостатки данного метода схожи с предыдущим, поскольку для обработки данных применяются CNN, но дополнительно требуется достаточно высокое качество данных.

Структура сети представлена на рисунке 3.

Применение стандартного подхода более предпочтительно, а также активно используется на рынке данных услуг. Применение иных методов также возможно, но требует высокого качества обучения используемых сетей. Также возможен и комплексный подход с применением двух методов, но существует необходимость в согласованной работе методов в продукте, иначе итоговый результат будет низкого качества.

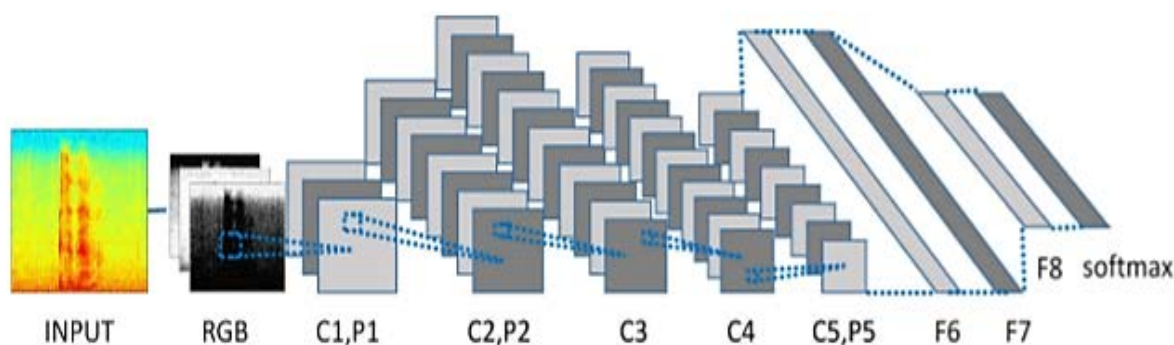


Рисунок 3 – Архитектура сети распознавания эмоций по голосу

В таблице 1 приведены достоинства и недостатки рассмотренных методов.

Таблица 1 – Сравнительный анализ алгоритмов распознавания эмоций

Метод	Достоинства	Недостатки
Классический метод	Проста по сравнению с иными методами, высокая точность	Зависимость от положения лица, освещения, качества изображения
Метод с применением глубокого обучения	Меньшая зависимость от положения лица, освещения, качества изображения	Сильная зависимость результатов от качества обучения
Метод оценки речи	Высокая точность при хорошем обучении, возможность применения в дополнение к иным методам	Сильная зависимость результатов от обучения, качества данных

Библиографические ссылки

1. Коровкин В. А. Современные методы распознавания эмоций // Молодежь и современные информационные технологии: сб. трудов XVI Международной научно-практической конференции студентов, аспирантов и молодых учёных, Томск, 3–7 дек. 2018 г. Томск : Изд-во ТПУ, 2019. С. 76–77.
2. Тимофеева О. П. Распознавание эмоций по изображению лица на основе глубоких нейронных сетей : [Электронный ресурс]. URL: <https://cyberleninka.ru/article/n/raspoznavanie-emotsiy-po-izobrazheniyu-litsa-na-osnove-glubokih-neyronnyh-setey> (дата доступа: 13.10.2021).
3. A breakthrough in Speech emotion recognition using Deep Retinal Convolution Neural Networks : [Электронный ресурс]. URL: <https://arxiv.org/ftp/arxiv/papers/1707/1707.09917.pdf> (дата доступа: 13.10.2021).
4. End-to-End Multimodal Emotion Recognition using Deep Neural Networks : [Электронный ресурс]. URL: <https://arxiv.org/pdf/1704.08619.pdf> (дата доступа: 13.10.2021).