УДК 004.032.26

О.П. Тимофеева, С.А. Неимущев, Л.И. Неимущева, И.А. Тихонов

РАСПОЗНАВАНИЕ ЭМОЦИЙ ПО ИЗОБРАЖЕНИЮ ЛИЦА НА ОСНОВЕ ГЛУБОКИХ НЕЙРОННЫХ СЕТЕЙ

Нижегородский государственный технический университет им. Р.Е. Алексеева

Рассматривается задача распознавания эмоций по изображению лица, полученному из видеопотока. Подход к решению основан на применении глубоких нейронных сетей. Приведен набор данных, используемый для обучения сети, его характеристики и распределение данных по классам эмоций. Описаны две модели сверточной нейронной сети: классическая сверточная нейронная сеть, построенная для данной задачи; сверточная нейронная сеть, улучшенная посредством механизмов регуляризации. На основе полученных результатов обучения сетей проведен сравнительный анализ точности классификации. Описан процесс распознавания эмоций на произвольных данных, не относящихся к рассматриваемому набору данных.

Ключевые слова: распознавание эмоций, классификация, машинное обучение, глубокое обучение, сверхточные нейронные сети, регуляризация.

Введение

Изучением эмоций и их проявления ученые занимаются достаточно давно. Ведь эмоции являются неизбежной частью любой межличностной коммуникации, выражают отношение человека к окружающему миру, сложившейся вокруг него ситуации, к самому себе. Вместе с тем, в последнее время потребность в выявлении человеческих эмоций еще более возросла. В первую очередь, это связано с расширением сферы применения задачи распознавания эмоций. В настоящее время это и мониторинг состояния водителя за рулем, и системы видео аналитики «умного города», и маркетинговые исследования, и системы безопасности.

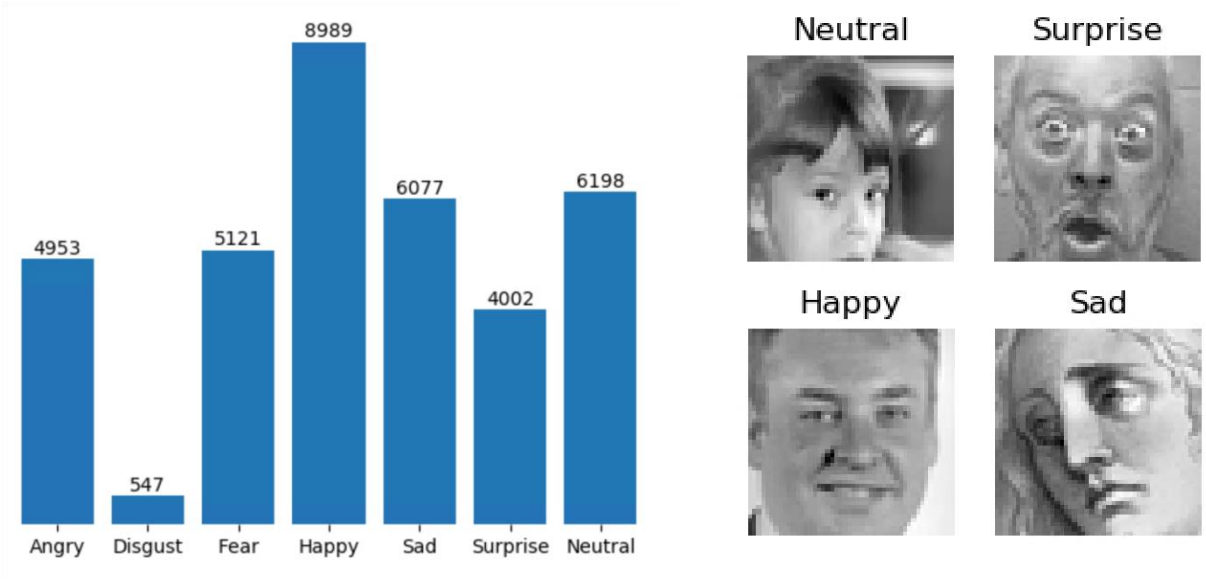
Эмоции могут быть выражены разными способами: мимикой, голосом, поведением, реакциями систем организма [1]. Наибольший интерес из них представляет распознавание эмоций человека по выражению его лица. Эта задача является достаточно популярной в настоящее время по ряду причин: такие изображения несложно получить, они содержат много полезной информации для распознавания эмоций, собрать большой набор данных в виде изображений лиц достаточно легко (по сравнению с другим материалом для распознавания: речью или образцами почерка).

Данная работа посвящена задаче распознавания эмоций по изображению лица человека. Для полного цикла исследования – формирования набора данных, создания, обучения и тестирования моделей использовался язык Python как один из наиболее популярных языков для решения задач в области анализа данных и машинного обучения.

Набор данных

В качестве набора данных для обучения глубоких сетей был выбран Facial Expression Recognition 2013 (FER2013), который был представлен на конференции International Conference on Machine Learning 2013 [2]. Этот набор данных содержит 35 887 изображений с разрешением 48×48пикселей, большинство из которых сделаны в произвольных условиях. База данных была создана с использованием инструментов поиска изображений Google. Каждое изображение классифицировано одним из семи видов эмоций: удивление (surprise), страх (fear), счастье(happy), гнев (angry), отвращение (disgust), грусть (sad) и нейтральное состояние или спокойствие (neutral). FER имеет большое число вариаций в изображениях, включая частичное закрытие лица (в основном, с помощью руки), низко контрастные изоб-

ражения и лица в очках. Распределение данных по разным классам эмоций и примеры изображений лиц с указанием классов, к которым они отнесены, представлены на рис. 1.



а) б)

Рис. 1. Набор данных FER 2013

а – диаграмма распределения данных по различным классам эмоций

б – примеры изображений лиц с указанием классов

В работе весь набор данных FER разделен на три части: обучающий набор, валидационный набор и тестовый набор. Первые два участвуют при обучении сети: обучающий набор используется для оптимизации весов модели, а валидационный набор предоставляет метрики после каждой эпохи обучения, которые помогают оценить качество обучения модели. Тестовый набор необходим для сравнения точности распознавания среди разных моделей.

Архитектура и особенности нейронных сетей

Для распознавания эмоций в работе используется архитектура сверточной нейронной сети (Convolutionalneuralnetwork, CNN). Схематично CNN представляет собой последовательность слоев. Каждый слой преобразует один активационный объем в другой с помощью дифференцируемой функции. Для организации сверточной нейронной сети применяется 3 основных слоя: свертка (convolution), пулинг (иначе слой подвыборки или субдискретизации, англ. pooling) и полносвязный (fullyconnected, FC) слой. Слои свертки и пулинга используются для извлечения карты признаков из исходного изображения, а полносвязные слои используются для конечной классификации изображения по извлеченным признакам.

Размер входного слоя сети равен 48×48×1, в соответствии с размером изображений из набора данных. Выходной слой сети – это вектор из 7 элементов, соответствующих вероятностям принадлежности входного изображения к каждому из классов. В результате входное изображение относится к классу, имеющему максимальное значение вероятности.

В процессе исследования были построены две модели CNN. Первая модель содержит 2 слоя свертки, 2 слоя подвыборки и 4 полносвязных слоя. Подробная иллюстрация первой модели: размерности слоев, их параметры и используемые функции активации представлены на рис. 2.

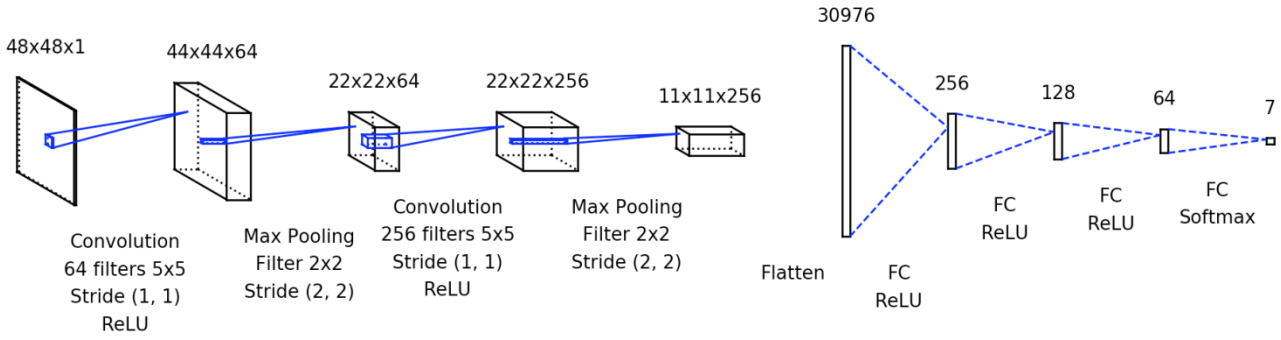


Рис. 2. Первая модель сверточной нейронной сети

Вторая модель является модернизацией первой модели и содержит 8 слоев свертки, 4 слоя подвыборки и 4 полносвязных слоя, а также механизм регуляризации [3]. Иллюстрация второй модели представлена на рис. 3.

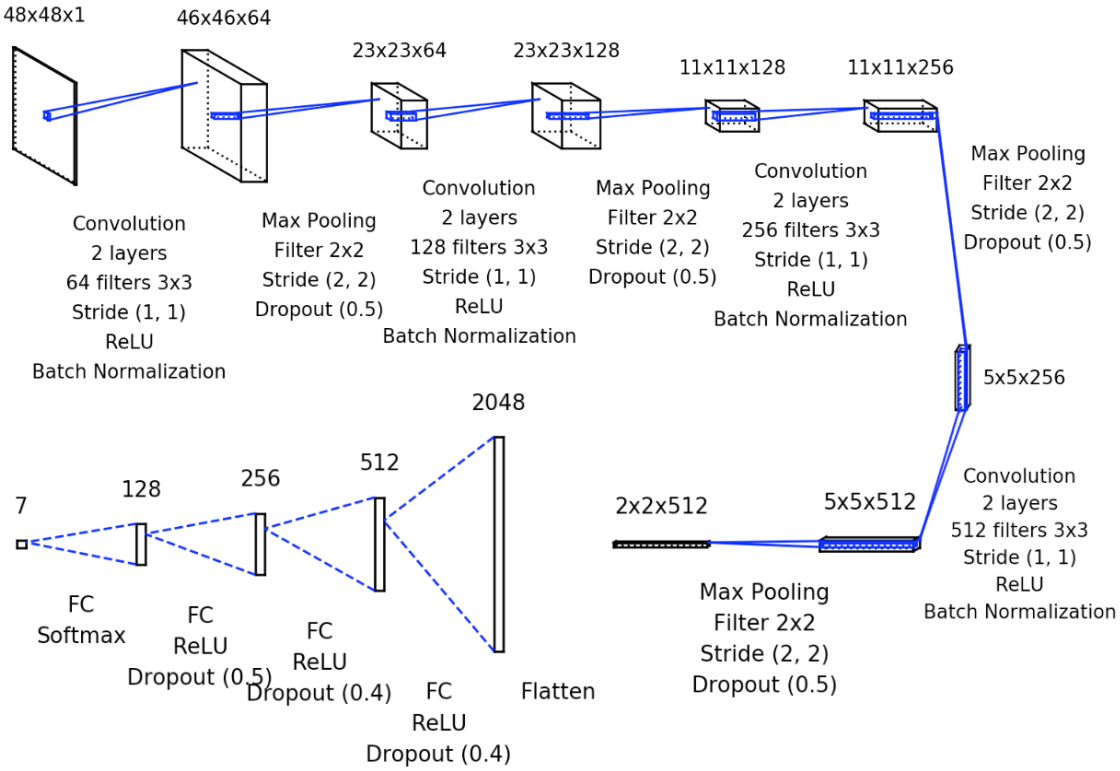


Рис. 3. Вторая модель сверточной нейронной сети

По сравнению с первой моделью, в ней большее количество сверточных слоев, которые имеют меньший размер матрицы свертки, что позволяет извлечь более детальную карту признаков. Механизм регуляризации позволяет избежать ситуации, называемой переобучением (overfitting) [4]. Характерным признаком переобучения является высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовой выборке. Такая ситуация может возникнуть, если данные имеют много признаков, но при этом сам набор данных содержит мало примеров, либо в том случае, когда модель является слишком сложной для данных. Во второй модели сети для предотвращения ситуации переобучения используются такие механизмы регуляризации, как Batch Normalization [5] и Dropout [6]. Рассмотрим идеи, лежащие в основе этих механизмов.

Обычно для обучения нейронной сети выполняется некоторая предварительная обработка входных данных. Например, набор данных FER нормализуется таким образом, чтобы его данные напоминали нормальное распределение – имели нулевое математическое ожидание и единичную дисперсию. Такая обработка происходит для предотвращения раннего насыщения нелинейных функций активации слоев и обеспечения того, чтобы все входные данные находились в одном диапазоне значений. Но проблема возникает в промежуточных слоях, поскольку распределение значений, которое может иметь активационная функция, постоянно меняется в процессе обучения. Это замедляет процесс обучения, потому что каждый слой должен учиться приспосабливаться к новому распределению на каждом этапе обучения. Эта проблема известна как внутренний ковариантный сдвиг.

Суть метода Batch Normalization заключается в нормализации входных значений внутренних слоев нейронной сети и, таким образом, предотвращении возникновения внутреннего ковариантного сдвига. В процессе обучения, механизм Batch Normalization выполняет следующие действия.

1. Вычисляется математическое ожиданиеμ_Bи дисперсияσ_B^2 входных значений слоя (1):

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i;$$
$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B). \quad (1)$$

2. Входные значения слоя нормализуются с помощью ранее рассчитанных статистических значений (2):

$$\overline{x_i} = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2}}. \quad (2)$$

3. Нормализованные значения масштабируются и сдвигаются для того, чтобы избежать изменения представления данных в слое (3):

$$y_i = \gamma \overline{x_i} + \beta. \quad (3)$$

При этом параметры масштабирования γ и сдвига β настраиваются во время обучения совместно с другими параметрами сети.

Основная идея механизма Dropout состоит в том, чтобы случайно отбрасывать отдельные нейроны в слоях (вместе с их связями) из нейронной сети во время обучения. Так как отброшенные нейроны перестают вносить свой вклад в процесс обучения сети, то это становится равносильно обучению новой нейронной сети. Это предотвращает слишком большую адаптацию нейронов друг к другу. Каждый слой, использующий Dropout, имеет параметр, определяющий вероятность исключения нейрона из сети.

Результаты исследования

После обучения первая сеть продемонстрировала точность распознавания эмоций 52 % на тестовом наборе данных. При этом на обучающем наборе данных точность распознавания составила 98 %. График изменения точности в процессе обучения модели представлен на рис. 4.

Матрица ошибок, построенная на тестовом наборе данных, представлена на рис. 5. В матрице ошибок строки и столбцы обозначены одним из семи классов эмоций. На пересечении указано количество вариантов, отнесенных к классу эмоций, обозначающему столбец, но реально принадлежащих к классу эмоций, обозначающему текущую строку. По матрице видно, что наименьшей ошибке подвержено распознавание эмоции «счастье» (23 % ошибок), наибольшей – распознавание эмоций «страх» и «гнев» (65 % ошибок).

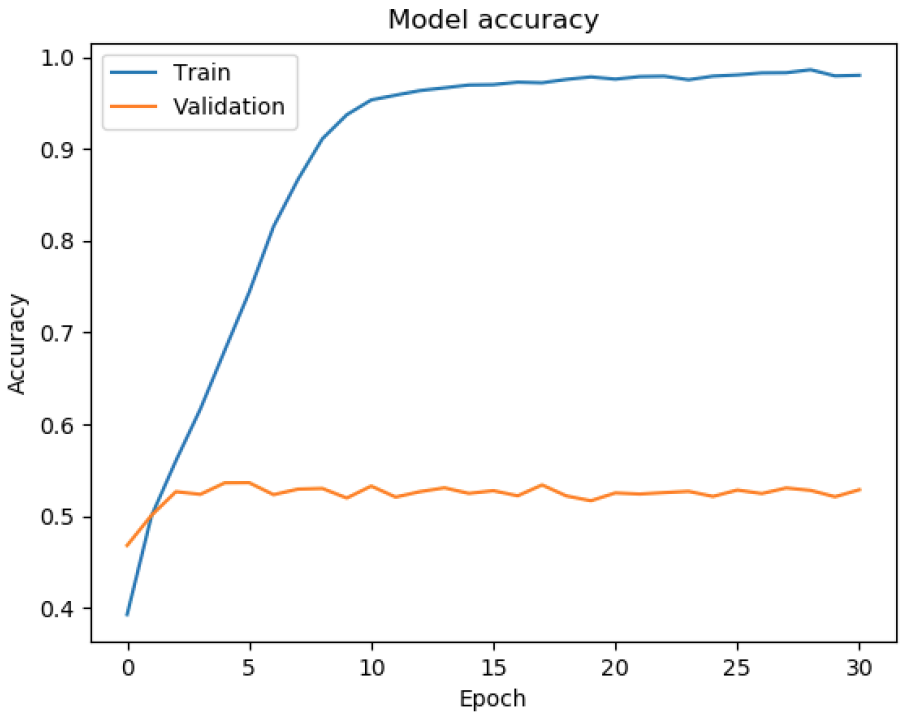


Рис. 4. График изменения точности распознавания модели в зависимости от эпохи обучения

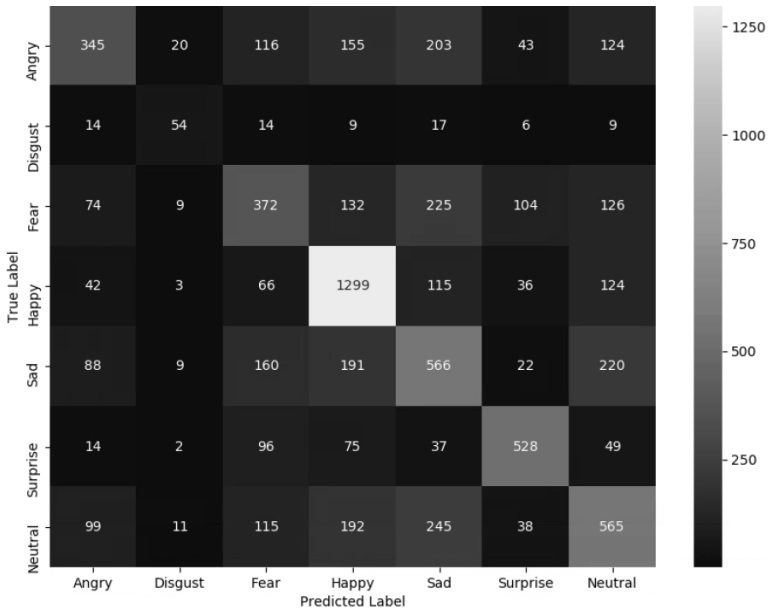


Рис. 5. Матрица ошибок первой модели

Для второй сверточной сети использовались тот же самый набор данных, что и для первой сети. В результате обучения сеть показывает точность распознавания 92 % на обучающем наборе данных, но при этом на валидационном наборе данных точность достигает 64 % (рис. 6). Корреляция между правильными и ошибочными распознаваниями представлена матрицей ошибок на рис. 7.

Высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовом наборе являются признаком переобучения сети. Как было указано ранее, решением данной проблемы является механизм регуляризации, который добавлен во вторую сверточную сеть.

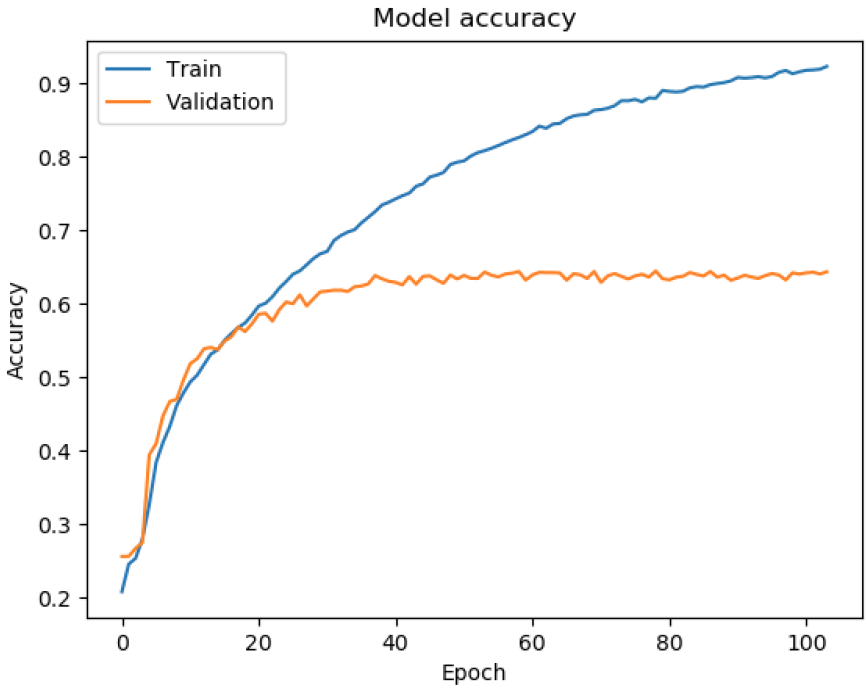


Рис. 6. График зависимости точности распознавания модели от номера эпохи обучения

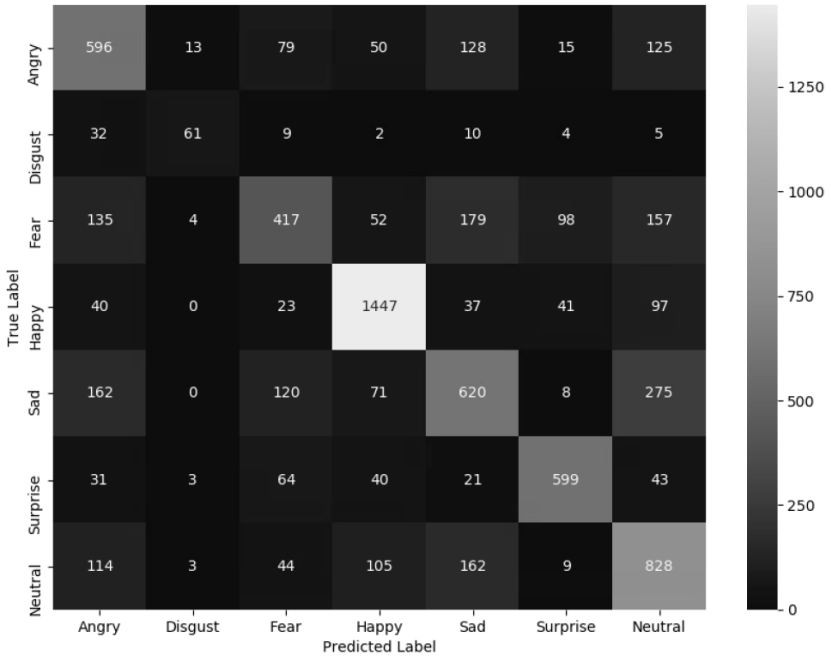


Рис. 7. Матрица ошибок второй модели

Анализ результатов позволяет сделать вывод, что регуляризация совместно с добавлением новых сверточных слоев в модели улучшила точность распознавания на 12 %.

Тестирование обученной модели на произвольных данных

Для тестирования обученной модели на произвольных данных было разработано вспомогательное приложение, которое позволяет классифицировать эмоции на заданном изображении или видео. В качестве источника данных может выступать как заранее записанное видео, так и видео, поступающее с камеры в реальном времени.

Для декодирования и покадровой обработки видео используется библиотека OpenCV [7]. Поиск лиц на отдельном кадре осуществляется методом Виолы-Джонса [8]. Данный метод демонстрирует высокую точность поиска лица на изображении вместе с быстрой скоростью работы. Также существуют альтернативные методы поиска лица, основанные на сверточных нейронных сетях, но они требуют большего количества ресурсов для обработки изображения [9], вследствие чего метод Виолы-Джонса является более приемлемым вариантом для классификации эмоций в реальном времени с высокой частотой кадров.

После выполнения поиска лиц по методу Виолы-Джонса все найденные лица на кадре подвергаются ряду преобразований для улучшения точности дальнейшей классификации.

1. Выравнивание положения лица по вертикали и горизонтали.
2. Гамма-коррекция [10].
3. Объединение нескольких цветовых каналов в один для получения изображения в градациях серого.
4. Изменение размера изображения до 48×48 пикселей.

Далее преобразованный набор лиц передается на вход классификатору – обученной модели. После завершения классификации каждому изображению лица будет присвоен соответствующий класс эмоции. Завершающим этапом обработки кадра является визуализация полученных классов – каждое найденное лицо на кадре обозначается цветной рамкой и маркируется названием присвоенной ему эмоции.

Схематично процесс обработки кадра изображен на рис. 8.

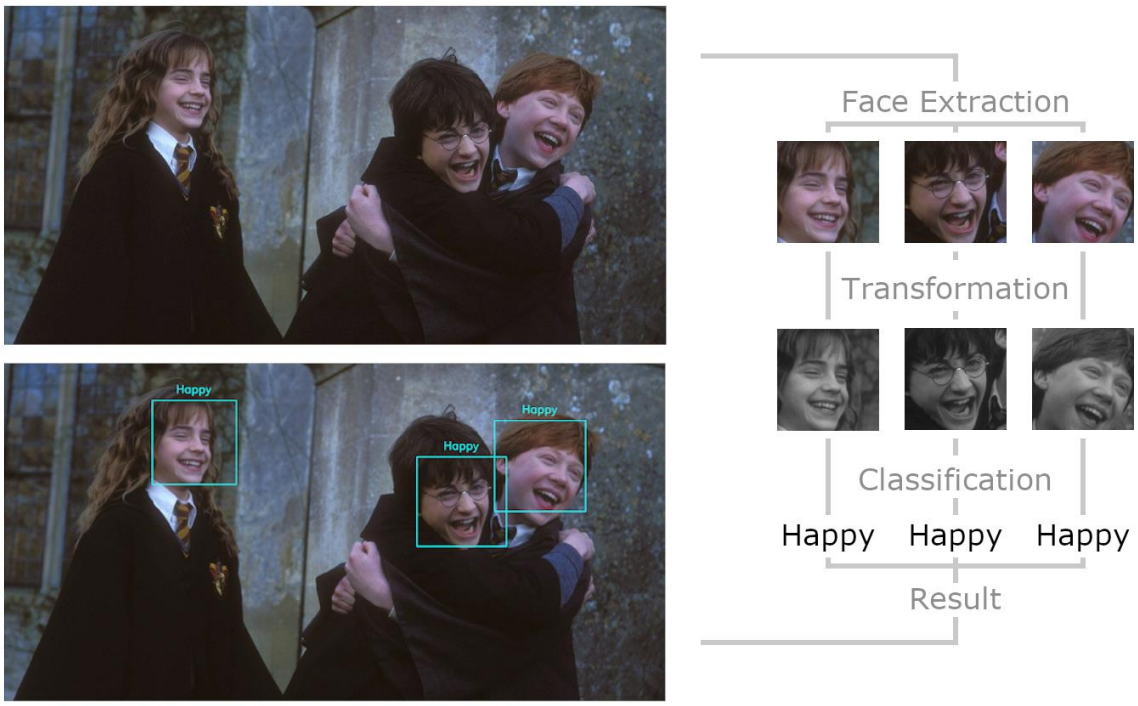


Рис. 8. Процесс обработки кадра

Заключение

В результате проведенного проектирования сетей и их последующего обучения наилучшая полученная точность классификации эмоций по изображению лица составила 64 %. При этом построенная матрица ошибок демонстрирует, что полученная точность классификации в первую очередь обусловлена неравномерным распределением данных по классам в исходном наборе данных. Так, количество изображений, отнесенных к классу «отвращение», в 16 раз меньше, чем количество изображений, отнесенных к классу «счастье».

Тестирование модели на произвольных данных, не относящихся к набору данных FER, позволило качественно оценить точность распознавания эмоций. Было выявлено, что из-за низкого разрешения входного изображения модели возникает погрешность в распознавании.

Дальнейшее исследование будет направлено как на улучшение используемого набора данных, так и на развитие текущей модели сверточной нейронной сети.

Библиографический список

1. **Gaind, B.** Emotion Detection and Analysis on Social Media / B. Gaind, V. Syal, S. Padgalwar // Global Journal of Engineering Science and Researches (ICRTCET-18). 2019. – P. 78-89.
2. Facial Expression Recognition Challenge// Deeplearning URL: http://deeplearning.net/icml2013-workshop-competition/challenges/ (датаобращения: 22.12.2019).
3. **Schmidhuber, J.** Deep Learning in Neural Networks: An Overview / J. Schmidhuber // Neural Networks. – 2015. – №61. – P. 85-117.
4. **Salman, S.** Overfitting Mechanism and Avoidance in Deep Neural Networks [Электронный ресурс] / S. Salman, X. Liu // arXiv.org. 2019. URL: https://arxiv.org/abs/1901.06566 (дата обращения: 17.01.2020).
5. **Ioffe, S.** Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [Электронный ресурс] / S. Ioffe, C. Szegedy // arXiv.org. 2015. URL: https://arxiv.org/abs/1502.03167 (дата обращения: 11.01.2020).
6. **Srivastava, N.** Dropout: A Simple Way to Prevent Neural Networks from Overfitting / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research. – 2014. – №15. – P. 1929-1958.
7. **Culjak, I.** A brief introduction to OpenCV / I. Culjak, D. Abram, T. Pribanic, H. Dzapo, M. Cifrek // 2012 Proceedings of the 35th International Convention MIPRO, Opatija. 2012. – P. 1725-1730.
8. **Viola, P.** Rapid object detection using a boosted cascade of simple features / P. Viola, M. Jones // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. – 2001. – T. 1.
9. **Murillo, P.C.U.** Comparison between CNN and Haar classifiers for surgical instrumentation classification / P.C.U. Murillo, R.J. Moreno, J.O.P. Arenas // Contemporary Engineering Sciences. – 2017. – T. 10. – № 28. – P. 1351-1363.
10. **Anila, S.** Preprocessing Technique for Face Recognition Applications under Varying Illumination Conditions / S. Anila, N. Devarajan // Global Journal of Computer Science and Technology Graphics & Vision. – 2012. – T. 12. – № 11.

Дата поступления
в редакцию: 02.02.2020

O.P. Timofeeva, S.A. Neimushchev, L.I. Neimushcheva, I.A. Tikhonov

FACIAL EMOTION RECOGNITION USING DEEP NEURAL NETWORKS

Nizhny Novgorod state technical university n.a. R.E. Alekseev

Purpose: The ability to recognize facial expressions automatically enables novel applications in human-computer interaction and other areas. This article is devoted to an approach to solving the problem of emotion recognition using deep learning networks.
Design/methodology/approach: Convolutional neural networks (CNN) are used for feature extraction and inference. Two different CNN architectures are proposed. As a training dataset, the FER2013 dataset is used.
Findings: The best achieved accuracy of emotion recognition on FER2013 dataset is 64%. Moreover, the obtained confusion matrices based on a test data set demonstrate classification problems caused by the uneven distribution of training dataset amongthe emotion classes.
Research limitations/implications: This research opens further prospects for both the development of the current CNN architecture and the expansion of the data set to address its shortcomings.
Originality/value: This approach can be used in applications requiring recognition of emotions from photos or videos.

Key words: emotion recognition, classification, machine learning, deep learning, convolutional neural networks, regularization.