ОТСЛЕЖИВАНИЕ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА С ПОМОЩЬЮ ТЕХНОЛОГИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Себякин А.С., Золотарюк А.В.

Финансовый университет при Правительстве Российской Федерации,
Россия, г. Москва, Ленинградский пр., д.49
sebyakin.a@ya.ru, azolotaryuk@fa.ru

Аннотация: В статье исследуется проблема распознавания эмоционального состояния человека по лицевым экспрессиям на статичном изображении. Представляется высокоточный, производительный авторский метод, позволяющий использовать его при экстремальных параметрах освещения, низком качестве изображения, частичных окклюзиях, при обработке больших массивов видео в реальном времени.

Ключевые слова: эмоции, распознавание эмоционального состояния, метод распознавания эмоций, нейронные сети, тренировка нейронной сети, ключевые точки лица, датасеты.

1 Актуальность

Эмоциональное состояние человека играет важную роль во многих сферах деятельности – при управлении критической инфраструктурой, в ходе медицинских операций, в различных областях бизнеса, при ведении военных действий, в процессах противодействия терроризму, в сфере обслуживания и т.п. Поэтому крайне важно отслеживать и контролировать данное состояние.

В банковской сфере данная технология может использоваться для оценивания адекватности и приветливости поведения сотрудников банка по отношению к клиентам, отслеживания общего настроения в коллективе, в скоринге персонала, для детекции конфликтов внутри коллектива или между сотрудником банка и клиентом, для определения степени удовлетворенности клиента.

Результаты рассмотренной работы, на взгляд авторов, обеспечат автоматизацию описываемых процессов, что делает проведенные исследования актуальными.

2 Решение

В настоящее время описано несколько моделей формы и особенностей человеческих лиц и признаков их классификации [1 - 10].

В данной статье предлагается использовать оригинальный подход к распознаванию эмоций по данному изображению на основе глубокой сверточной нейросети в качестве классификатора.

Процесс тренировки нейронной сети состоит из четырех стадий:

- детекция лица на изображении, его локализация;
- аугментация изображения лица;
- нормализация изображения;
- обработка изображения нейросетью и обратное распространение ошибки.

Рассмотрим указанные стадии более подробно.

Детекция и локализация лица выполнялась с помощью алгоритма FaceBoxes, но может быть выполнена любым другим доступным методом, таким как YOLO, SSD, Haar Cascade, HOG+SVM и т.п.

Аугментация изображения лица подразумевает изменения в изображении, такие как добавление шума, повороты, отзеркаливание, изменение яркости и контраста, блюр. При использовании нейросети после тренировки аугментация не производится. Использование аугментаций позволяет добиться значительного увеличения точности в условиях экстремального освещения, частичных окклюзиях и общего низкого качества входного изображения.

Нормализация изображения представляет собой деление значений каждого пикселя на 255, переводя их в интервал [0, 1].

Обработка изображения нейросетью и обратное распространение ошибки выполнялось на видеокарте nvidia geforce gtx 1080 с помощью фреймворка pytorch.

Для тренировки нейросети был использован оптимизатор Adam с начальной скоростью обучения 2e-4. Далее скорость обучения уменьшалась в 2 раза после 10 эпох без увеличения точности на валидации. Размерность входного тензора – (32, 256, 256). Функция потерь: Cross Entropy с Softmax[112].

Обучение нейросети выполнялось на датасете AffectNet, включающем один миллион изображений с лицами, собранными из интернета с помощью поисковых систем по 1250 ключевым словам на шести языках. Вручную размечены 420 тысяч изображений с указанием базовых эмоций (злоба, отвращение, страх, радость, грусть, удивление, без эмоций) и значений валентности и возбужденности.

В качестве базовой архитектуры нейросети была выбрана архитектура ResNet по причине ее высокой точности в задачах классификации, широкого распространения, обилия оптимизаций на уровне фреймворка и низкоуровневых оптимизаций на видеокарте (cuDNN).

3 Критерии оценки

Для оценки качества предсказаний нейросети использовались метрики top-1 accuracy, top-3 accuracy. Их графики изображены на рис. 1.

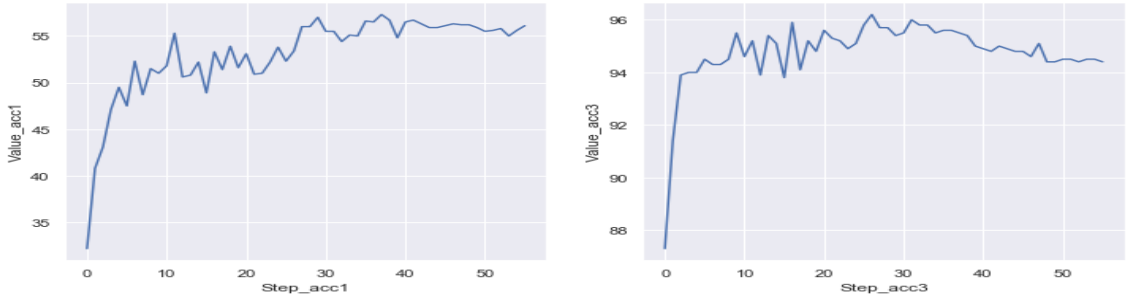

Рис. 1. Графики top-1 accuracy (слева), top-3 accuracy (справа). По оси x – количество пройденных эпох (полных итераций по датасету)

[112] *Википедия.URL: https://en.wikipedia.org/wiki/Cross_entropy (дата обращения: 24.06.2019); URL: https://en.wikipedia.org/wiki/Softmax_function (дата обращения: 24.06.2019)*

Top-K Accuracy – это такая вероятность, с которой правильный ответ содержится среди первых K предсказаний алгоритма в общем списке предсказаний, отсортированных по параметру confidence, т.е «уверенности» алгоритма в правильности ответа.

Максимальная точность модели на валидации составила 57.3% (top-1 accuracy) и 96.2% (top-3 accuracy). Матрица несоответствия (Confusion Matrix) представлена на рис. 2.

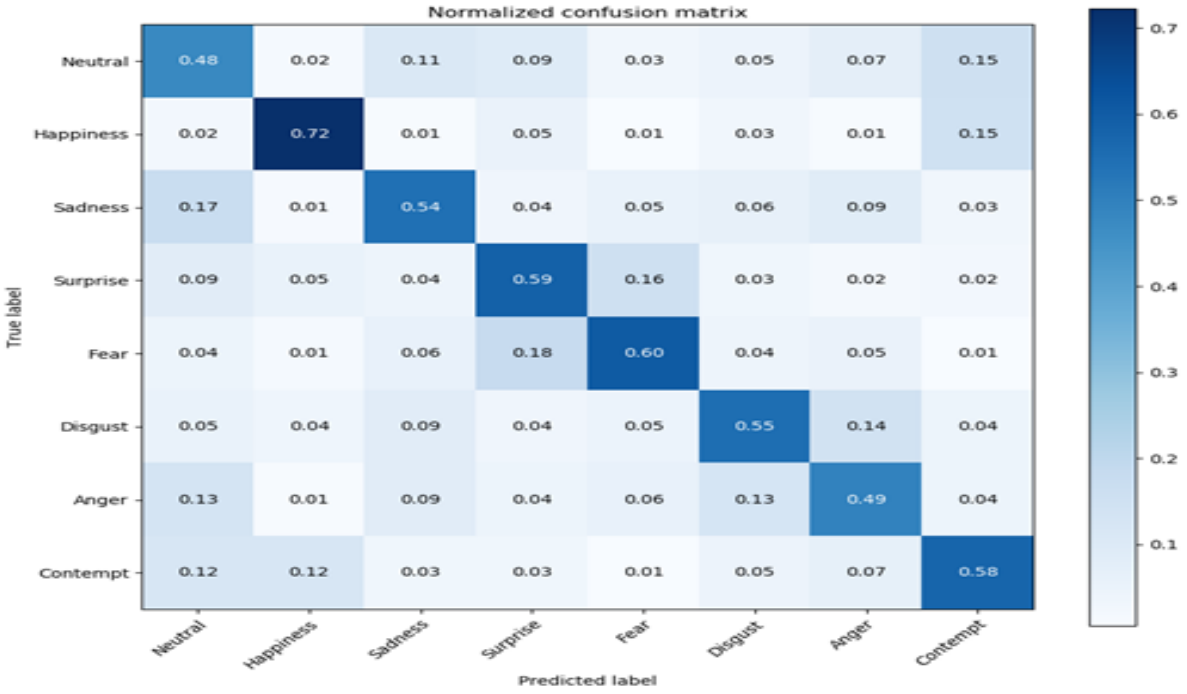

Рис. 2. Матрица несоответсвия на валидационной части датасета

Выводы

Проведенные эксперименты подтверждают эффективность предлагаемого подхода с применением интеллектуальных технологий нейросетей для распознавания лицевых экспрессий на статичном изображении.

Дальнейшие исследования планируется проводить в направлении адаптации данного метода для распознавания лицевых экспрессий на видео, используя не только признаки и зависимости внутри изображения (spatial features), но и признаки между кадрами, проявляемые во времени (temporal features).

Полученные результаты, как представляется, найдут применение в различных системах распознавания поведения людей и предсказания их дальнейших действий, что является немаловажным и актуальным во многих сферах деятельности человека, в том числе в банковском деле.

Литература

1.*Mollahosseini, A., Hasani, B., & Mahoor, M. H.* (2017). “Affectnet: A database for facial expression, valence, and arousal computing in the wild”. IEEE Transactions on Affective Computing.

2. *Zhang, S., Zhu, X., Lei, Z., Shi, H., Wang, X., & Li, S. Z.* (2017, October). “Faceboxes: A CPU real-time face detector with high accuracy”. In 2017 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-9). IEEE.

3. *Li, Ying & Brown, Lisa & Hampapur, Arun & Pankanti, S & Senior, Andrew & Bolle, Ruud.* (2019). “Real World Real-time Automatic Recognition of Facial Expressions”. Exploratory Computer Vision Group, IBM T. J. Watson Research Center, pp.1-8.

4. *H.-W. Ng, V. D. Nguyen, V. Vonikakis, and S. Winkler*, “Deep learning for emotion recognition on small datasets using transfer learning,” in Proceedings of the 2015 ACM on international conference on multimodal interaction. ACM, 2015, pp. 443–449.

5. *Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong*, “Identity-aware convolutional neural network for facial expression recognition,” in Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference on. IEEE, 2017, pp. 558–565.

6. *M. Shin, M. Kim, and D.-S. Kwon*, “Baseline cnn structure analysis for facial expression recognition,” in Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on. IEEE, 2016, pp. 724–729.

7. *Y. Sun, X. Wang, and X. Tang*, “Deep convolutional network cascade for facial point detection,” in Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. IEEE, 2013, pp. 3476–3483.
8. *S. Ren, X. Cao, Y. Wei and J. Sun*, "Face Alignment at 3000 FPS via Regressing Local Binary Features," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, 2014, pp. 1685-1692.
9. *Schroff, F., Kalenichenko, D., & Philbin, J.* (2015). Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 815-823.
10. *He, K., Zhang, X., Ren, S., & Sun, J.* (2016). Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770-778.