

Аликрицкий Михаил Сергеевич, магистрант,
Поволжский государственный университет
телекоммуникаций и информатики

РАСПОЗНАВАНИЕ ЭМОЦИЙ ПО ВЫРАЖЕНИЮ ЛИЦА НА НАБОРЕ ДАНЫХ FER-2013 С ИСПОЛЬЗОВАНИЕМ НЕЙРОННОЙ СЕТИ

Аннотация. В работе исследуется применение сверточных нейронных сетей (CNN) для классификации эмоций по мимике лица. Эксперименты проводятся на общедоступном наборе данных FER-2013, содержащем более 35 тысяч изображений с семью метками эмоций. Предложенная архитектура CNN демонстрирует устойчивость к шумам и низкому разрешению изображений, характерным для реальных условий. Достигнутая точность распознавания сопоставима с лучшими решениями на данном датасете.

Ключевые слова: Распознавание эмоций, выражение лица, набор данных FER-2013, сверточная нейронная сеть, глубокое обучение, компьютерное зрение, аугментация данных.

Выражения лица играют очень важную роль в человеческом общении, поскольку они передают эмоции, намерения, а также другие скрытые социальные сигналы. Технологии автоматического распознавания эмоций по выражению лица (FER) развиваются, чтобы извлекать эти эмоциональные состояния из изображений или видеозаписей и находить применение в интеллектуальных обучающих системах, человеко-компьютерных интерфейсах, учитывающих эмоции, психическом здоровье, социальной робототехнике, видеонаблюдении и анализе поведения клиентов. Надежная система FER способна повысить качество взаимодействия, делая технологию более чувствительной к человеческим эмоциям.

Даже при значительном прогрессе в глубоком обучении сложно создать системы FER, которые могли бы работать в неконтролируемых условиях реальной жизни. Захваченные и сфотографированные изображения обычно не имеют высокого разрешения, искажены освещением, есть окклюзии, различия в положении головы и крайне тонкие изменения в выражении лица. Кроме того, метки эмоций по своей природе субъективны, а наборы данных обычно содержат спорадические и зашумленные аннотации. Ситуация усугубляется также значительной диспропорцией в классах, где некоторые эмоции, такие как отвращение или страх, представлены меньше, чем нейтральные или счастливое лицо. В совокупности эти переменные оказывают негативное влияние на обобщающую способность и могут использоваться для получения предвзятых прогнозов.

Набор данных FER-2013, используется в качестве одного из наиболее часто упоминаемых эталонных наборов для распознавания эмоций по лицу. Он состоит примерно из 35000 изображений лиц в оттенках серого размером 48x48 и сгруппирован в семь категорий эмоций. Несмотря на свою популярность, этот набор данных известен своей сложностью, поскольку качество изображений и меток низкое, и наблюдается явный дисбаланс между классами.

FER-2013 содержит 35887 изображений, размеченных по семи категориям: гнев, отвращение, страх, радость, нейтральное состояние, грусть, удивление. Разделение данных стандартное, 87,5 процентов обучающий набор и 12,5 процентов тестовых. Во время обучения применялась аугментация в реальном времени (вращение, сдвиг по ширине/высоте, масштабирование, сдвиг, горизонтальное отражение) с помощью Keras ImageDataGenerator. Была использована предобученная модель EfficientNetB2, с добавлением головы в виде GAP, dropout (0.5) и dense (7).



На рисунке 1 показана динамика обучения предложенной модели FER на основе EfficientNetB2 в течение 10 эпох, состоящих из 3 эпох фазы разогрева, за которой следует 7 эпох фазы тонкой настройки.

На рисунке 1 показана эволюция точности обучения и валидации. В течение начальных эпох точность обучения быстро возрастает, поскольку классификационный модуль адаптируется к распределению эмоций FER-2013, в то время как базовая модель остается неизменной. Точность валидации демонстрирует аналогичную закономерность, что свидетельствует о хорошей работе признаков, перенесенных из предварительно обученных весов ImageNet.

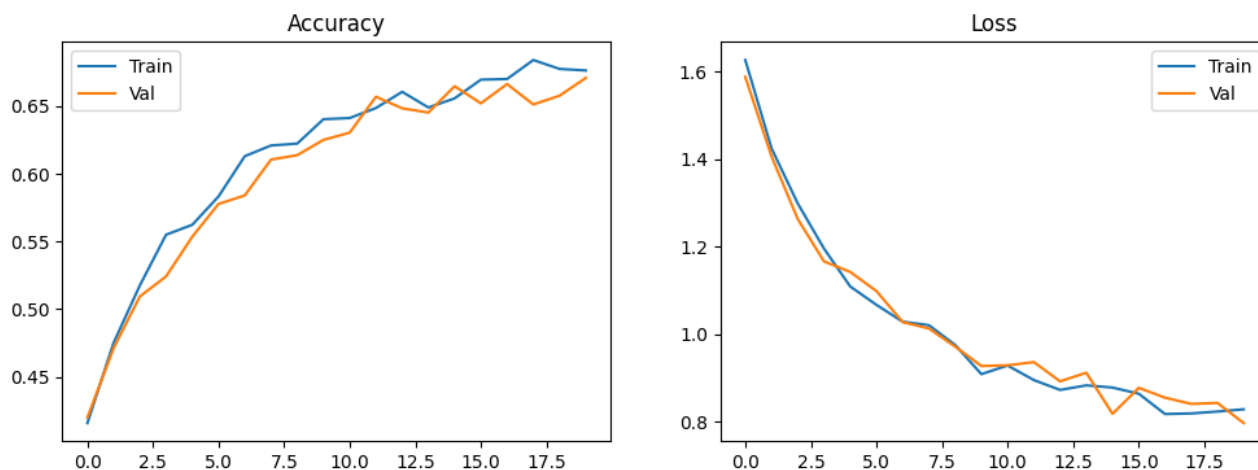


Рисунок 1. Графики обучения

Почти полное совпадение кривых обучения и валидации указывает на то, что используемые стратегии регуляризации, а именно сглаживание меток, дропаут и широкомасштабное расширение данных, эффективны в предотвращении переобучения и обеспечивают стабильное и надежное обучение. Достижимая точность 0.6536. В таблице 1 показаны метрики.

Таблица 1.

Сравнение точности классификации различных методов

Class	Precision	Recall	F1
Angry	0.54	0.60	0.57
Disgust	0.45	0.71	0.55
Fear	0.53	0.38	0.44
Happy	0.88	0.83	0.85
Neutral	0.56	0.67	0.61
Sad	0.55	0.48	0.51
Surprise	0.72	0.83	0.77
Overall	0.65	0.65	0.64

Высокие значения F1-меры для радости и удивления указывают на высокую способность распознавания визуально различимых выражений. Более низкие показатели для



страха и грусти отражают присущую этим эмоциям неоднозначность и визуальное сходство, что является известным ограничением FER-2013.

Предложенное решение на основе EfficientNetB2 с правильно подобранными методами регуляризации и аугментации достигает приемлемой для бенчмарка FER-2013 точности (65,4%), однако фундаментальные ограничения самого датасета (низкое качество, дисбаланс, нечёткость меток) остаются главным препятствием на пути к созданию универсальной системы распознавания эмоций в диких условиях. Дальнейшие исследования должны быть направлены на комбинирование нескольких датасетов и использование контекстной информации.

Список литературы:

1. Ekman P. Emotions Revealed: Recognizing Faces and Feelings / P. Ekman. – New York: Times Books, 2003. – 304 p.
2. Dalal N. Histograms of Oriented Gradients for Human Detection / N. Dalal, B. Triggs // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2005. – Vol. 1. – P. 886–893.
3. Ojala T. Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns / T. Ojala, M. Pietikäinen, T. Mäenpää // IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2002. – Vol. 24, No. 7. – P. 971–987.
4. Bishop C.M. Pattern Recognition and Machine Learning / C.M. Bishop. – New York: Springer, 2006. – 738 p.
5. Goodfellow I. Deep Learning / I. Goodfellow, Y. Bengio, A. Courville. – Cambridge: MIT Press, 2016. – 775 p.
6. Calvo R.A. Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications / R.A. Calvo, S. D’Mello // IEEE Transactions on Affective Computing. – 2010. – Vol. 1, No. 1. – P. 18–37.
7. Shan C. Facial Expression Recognition Based on Local Binary Patterns: A Comprehensive Study / C. Shan, S. Gong, P.W. McOwan // Image and Vision Computing. – 2009. – Vol. 27, No. 6. – P. 803–816.
8. Happy S.L. Automatic Facial Expression Recognition Using Features of Salient Facial Patches / S.L. Happy, A. Routray // IEEE Transactions on Affective Computing. – 2015. – Vol. 6, No. 1. – P. 1–12.
9. Li S. Deep Facial Expression Recognition: A Survey / S. Li, W. Deng // IEEE Transactions on Affective Computing. – 2020. – Vol. 13, No. 3. – P. 1195–1215.

