УДК 004.032.26

РЕШЕНИЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ЭМОЦИЙ ЛЮДЕЙ С ПОМОЩЬЮ АРХИТЕКТУР RESNET, EFFICIENTNET, CONVNEXT

В.А. Афанасьев[1]

[1]Самарский университет, г. Самара
e-mail: listussr85@gmail.com

Аннотация

В работе исследуется решение задачи распознавания эмоций людей по лицевым выражениям с помощью сверточных нейронных сетей. Рассматривались архитектуры семейств ResNet, EfficientNet, ConvNeXt. Показано, что рассмотренные архитектуры могут показывать качественные результаты, несмотря на дисбаланс классов эмоций.

Ключевые слова: сверточные нейронные сети, классификация, f1-мера, тепловые карты градиентов, распознавание эмоций, матрицы ошибок.

Введение

Задача распознавания эмоций людей по изображениям лиц – одна из актуальных задач компьютерного зрения. На данный момент задача имеет различные решения, начиная от классических геометрических алгоритмов, заканчивая нейронными сетями.

В работе решается вопрос применимости к задаче распознавания эмоций таких семейств свёрточных сетей, как ResNet [1], EfficientNet [2], ConvNeXt [3].

1. Исходные данные и рассматриваемые архитектуры

В ходе исследования используется уменьшенная версия AffectNet [4], в которой содержится 28000 изображений. Все изображения находятся в формате RGB и имеют размер 96×96 пикселей. Для решения задачи были взяты сети ResNet-50, EfficientNet-B3 и ConvNeXt-Tiny.

Все сети инициализировались весами ImageNet-1k и обучались в 2 этапа: обучение только классификатора и дообучение всей сети. В обучении использовалась ранняя остановка по f1-мере с выдержкой в 3-5 эпох. Основные гиперпараметры приведены в таблице 1.

Таблица 1 – Основные гиперпараметры при обучении сетей

Параметр	ResNet-50	EfficientNet-B3	ConvNeXt_tiny
Темп обучения слоёв	$1 * 10^{-5} - 5 * 10^{-5}$	$3 * 10^{-5}$	$5 * 10^{-5}$
L2 Регуляризация	$5 * 10^{-2}$	$1 * 10^{-4}$	$5 * 10^{-2}$
Оптимизатор	Adam	Adam	AdamW
Планировщик	С косинусным отжигом	С косинусным отжигом	С косинусным отжигом

2. Результаты обучения

ResNet-50 даёт хорошие результаты на классах нейтральности, счастья, грусти и страха – их полнота выше 0,75, как показано на рисунке 1. Однако классы гнева и отвращения имеют полноту ниже 0,5. Общая средняя f1-мера составила 0,6686. Из карты градиентов видно, что для гнева модель фокусируется на носу и переносице, хотя основные маркеры находятся выше – это брови и глаза. У отвращения и удивления ситуация похожа – модель пропускает признаки, связанные с глазами.

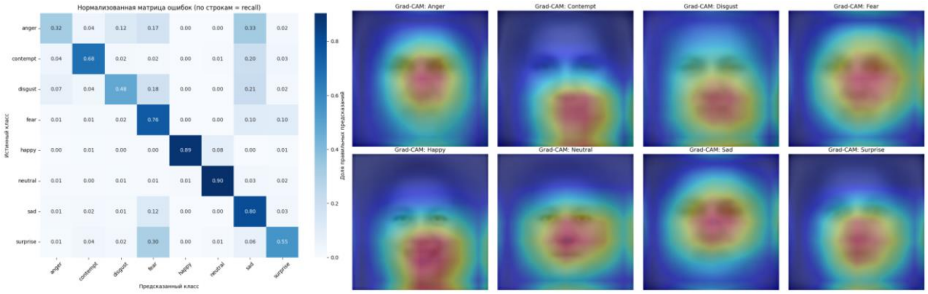


Рисунок 1 – Тепловая карта градиентов ResNet-50

Модификация ResNet-50 со слоями пространственного и канального внимания показала чуть лучшие результаты средней f1-меры – 0,6714. Все отстающие классы теперь имеют полноту выше 0,5, однако ухудшились классы страха и грусти. Причина видна на тепловой карте на рисунке 2. У страха внимание больше сместилось на глаза, пропуская открытый рот. У грусти модель стала меньше фокусироваться на глазах.

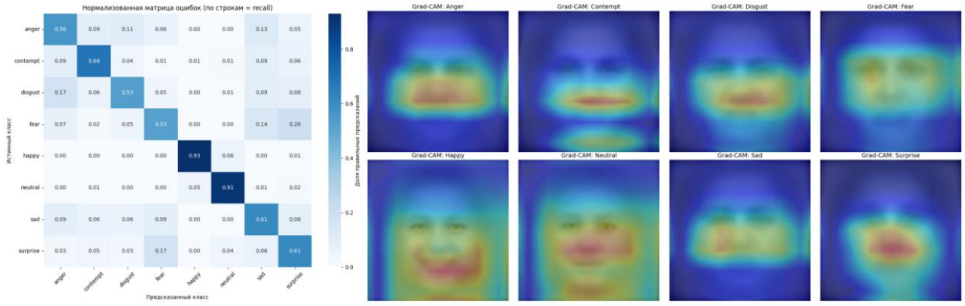


Рисунок 2 – Тепловая карта градиентов ResNet-50 с вниманием

EfficientNet-B3 показала более высокую среднюю f1-меру – 0,6745. Почти все классы имеют полноту выше 0,55 за исключением страха. Но появилась проблема – градиенты «грусти» почти равномерно рассредоточены по изображению с пиками интенсивности на подбородке и в верхнем правом углу изображения. Сеть не научилась выделять достаточно характерных признаков для этого класса.

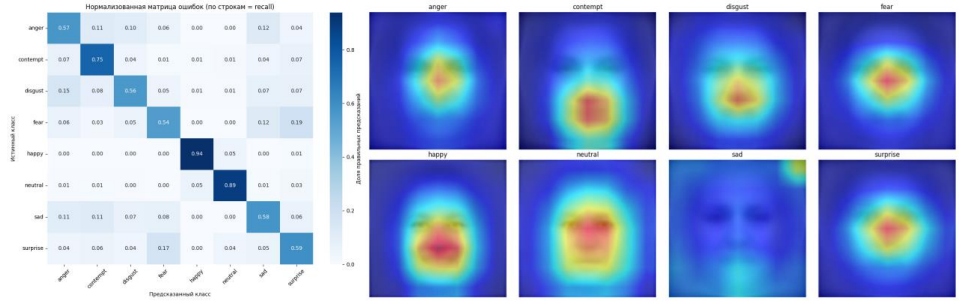


Рисунок 3 – Матрица ошибок и тепловая карта градиентов EfficientNet-B3

Для EfficientNet-B3 был также проведен эксперимент с добавлением внимания, однако здесь была произведена замена глобального объединения с усреднением на глобальное объединение с вниманием. Модификация сильно ухудшила предсказания сети – средняя f1-мера упала до 0,6476. Сеть перестала видеть ключевые признаки и стала охватывать вниманием всё лицо.

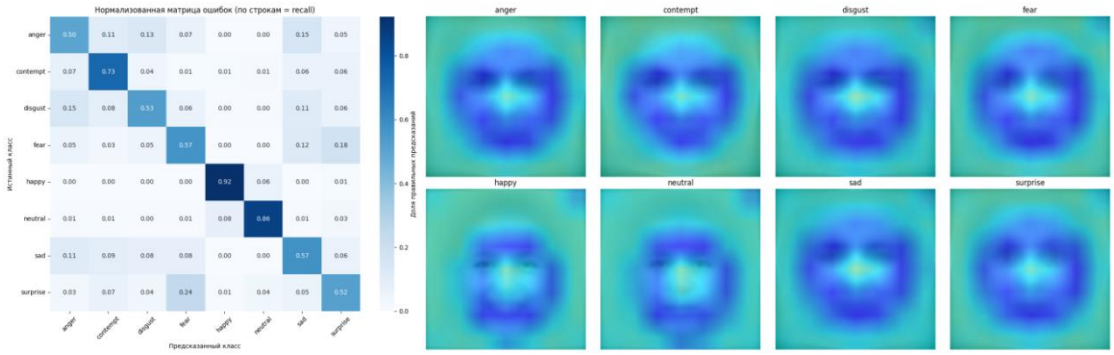


Рисунок 4 – Матрица ошибок и тепловая карта EfficientNet-B3 с вниманием

ConvNeXt оказалась лучшей моделью в плане средней f1-меры. Значение метрики составило 0,7006. Нейронная сеть стала хорошо предсказывать отвращение, презрение и удивление, однако имеет проблему – класс грусти с полнотой 0,5. На тепловой карте видно градиенты, описывающие анатомически правильные признаки эмоций. Так для страха и удивления внимание сети сконцентрировано на области глаз и бровей с небольшими пиками в области рта, а у отвращения и презрения внимание сконцентрировано чуть выше рта, захватывая характерные морщины и уголки губ.

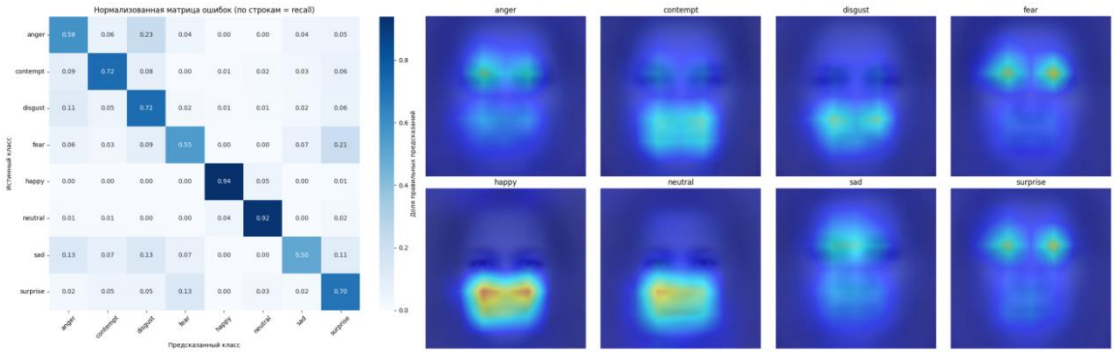


Рисунок 5 – Матрица ошибок и тепловая карта градиентов ConvNeXt_tiny

Таблица 2 – Результаты рассматриваемых архитектур на тестовой выборке

Название сети	**Точность**	**f1-мера**
ResNet-50	0,7027	0,6686
ResNet-50 + внимание	0,7050	0,6714
EfficientNet-B3	0,7072	0,6745
EfficientNet-B3 + внимание	0,6805	0,6476
ConvNeXt_tiny	0,7330	0,7006

Заключение

Эксперименты показывают, что можно добиться высоких результатов используя свёрточные архитектуры. Самыми стабильными сетями оказались ResNet-50 со слоями внимания и ConvNeXt_tiny. Худшие результаты показали ResNet-50 без модификаций и EfficientNet-B3 с вниманием. У ResNet есть проблемные классы, которые сеть почти не распознаёт, а EfficientNet с вниманием имеет более низкое качество, что видно из таблицы 2. Наиболее применимой к задаче оказалась сеть ConvNeXt_tiny.

Однако есть классы, на которых есть проблемы у всех архитектур – это гнев, грусть и страх. Причиной является сложность улавливания микровыражений с учётом малого размера изображений и сильного дисбаланса классов в датасете.

Дальнейшим направлением работы будет являться приложение полученных алгоритмов к распознаванию эмоций по видеопотоку и к улучшению стабильности классификации в случае работы с зашумлёнными данными.

Список использованных источников

1. Deep Learning review of ResNet architecture for lung disease identification in CXR image / S. Hasanah, A. Pravitasari, A. Abdullah [et al.] // Applied Sciences. – 2023. – V. 13.

2. EfficientNet for human FER using transfer learning / R. Singh, H. Sharma, N. K. Mehta [et al.] // Indian Journal of Science and Computing. – 2022. – P. 2792.

3. Nan, B. A Visual self-attention mechanism facial expression recognition network beyond convnext / B. Nan, F. Liu, X. Qian, W. Song // arXiv. – 2025. – arXiv:2504.09077v1.

4. Kim, D. Label quality in AffectNet: results of crowd-based re-annotation / D. Kim, C. Wallraven // Lecture Notes in Computer Science. – 2022. – V. 13189. – P. 518–533.

CONFERENCE PROCEEDINGS GUIDELINES

V.A. Afanasiev[1]

[1]Samara University, Samara
e-mail: listussr85@gmail.com

Abstract

The paper explores the solution of the problem of recognizing people's emotions by facial expressions using convolutional neural networks. Architectures of the ResNet, EfficientNet, and ConvNeXt families were considered. It is shown that the considered architectures can show high-quality results, despite the imbalance of emotion classes.

Keywords: convolutional neural networks, classification, f1-measure, gradient heat maps, emotion recognition, confusion matrices.