

НЕЙРОСЕТЕВОЕ РАСПОЗНАВАНИЕ ЭМОЦИЙ НА СТАТИЧНЫХ ИЗОБРАЖЕНИЯХ

П.А. Кочеткова, Е.И. Козлова

*Белорусский государственный университет, Минск, Беларусь
kochetkova1303@gmail.com*

Представлена уточненная модель выражения эмоций на лице человека для их распознавания по изображению. Для реализации процесса распознавания использована сверточная нейронная сеть. Целевой переменной является эмоция, которая может принимать одно из значений удивление, счастье, гнев, печаль. Для увеличения количества изображений в обучающей выборке применено аугментирование данных наложением гауссовского шума. Исследования показали, что после уточнения модели количество правильно распознанных эмоций на изображениях повысилось на 5 - 12 % в зависимости от типа эмоции.

Ключевые слова: распознавание эмоций; сверточная нейронная сеть; набор данных; аугментация.

ВВЕДЕНИЕ

Распознавание эмоций по изображениям лица человека является сегодня актуальной задачей при создании приложений и алгоритмов управления в различных областях человеческой деятельности – рекламном деле, социальной и медицинской сферах, образовании, технических областях, на транспорте, в развлекательном бизнесе и пр. В последнее время эта задача решается с помощью нейросетевых алгоритмов [1]. В ряде работ исследователей нейросетевых методов распознавания эмоций [2–4] описаны разные архитектуры свёрточных нейронных сетей, применяемые для распознавания эмоций на изображениях, полученных из видеопотока. Чаще всего исследователи применяют сверточные нейронные сети как ставшие уже классическими, так и с изменением числа слоев, функций активации и пр. Так, авторы [2] указали, что при использовании двух разных архитектур после обучения полученная в первом случае точность распознавания эмоций составила всего 52 % на тестовом наборе данных, а после изменения архитектуры – 64 %. При этом на обучающем наборе данных точность распознавания составила в обоих случаях около 98 %. В данной работе исследованы возможности повышения качества распознавания эмоций путем уточнения моделей описания некоторых базовых эмоций человека и применения сверточной нейронной сети для их распознавания в соответствии с уточненными моделями.

ИНСТРУМЕНТЫ ИССЛЕДОВАНИЯ

Сверточные нейронные сети (CNN) схематично представляют собой последовательность слоев, каждый из которых преобразует один активационный объем в другой с помощью дифференцируемой функции. Для организации сверточной нейронной сети применяется 3 основных слоя: свертка, пулинг и полносвязный слой. Слои свертки и пулинга используются для извлечения карты признаков из исходного изображения, а полносвязные слои используются для конечной классификации изображения по извлеченным признакам [5].

Для проведения исследований по распознаванию эмоций на изображениях лиц в качестве инструмента были выбраны свёрточные нейронные сети (CNN). Выбранная архитектура CNN содержит 2 слоя свертки, 2 pooling-слоя и 4 полносвязных слоя. Архитектура подобрана так, чтобы результаты можно было сравнить с результатами [2]. Отличием был размер входного слоя сети (он равнялся $256 \times 256 \times 1$), определенный в соответствии с размером изображений из набора данных. Выходной слой сети – это вектор из 4-х компонентов со значениями от 0 до 1, указывающими степень принадлежности предъявленной на изображении эмоции к выбранному классу. В данном исследовании также применена функция активации ReLU. Для обучения применен метод обучения с учителем, при котором сеть обучается на большом наборе аннотированных изображений лиц с различными эмоциями. Для улучшения производительности и предотвращения переобучения применен метод градиентного спуска.

УТОЧНЕНИЕ МОДЕЛЕЙ БАЗОВЫХ ЭМОЦИЙ И ПОДГОТОВКА ИЗОБРАЖЕНИЙ

К базовым эмоциям принято относить печаль, гнев, удивление, страх, отвращение, презрение, радость (эмоции удовольствия). В данной работе рассматривались следующие эмоции: удивление, счастье (радость), гнев, печаль [6].

Для обучения нейронной сети необходимо иметь большой набор данных, содержащий изображения лиц с различными эмоциями. В качестве основного набора данных для исследований был использован "Facial Emotion Recognition Image Dataset" [7]. Выбранный датасет содержит изображения людей разных возрастов, пола, расы, разной степени экспрессии выражения эмоций.

Анализ результатов [2] позволяет сделать предположение, что высокая точность распознавания на обучающей выборке и относительно низкая точность распознавания на тестовом наборе являются признаком переобучения сети. При этом регуляризация совместно с добавлением новых сверточных слоев в модели улучшила точность распознавания на 12% [2].

Мы предположили, что изменение моделей базовых эмоций и отбор изображений для обучения нейронной сети согласно уточненным признакам модели, позволит улучшить качество распознавания эмоций по изображению лица человека. Для проверки этого предположения и с целью уменьшить неоднозначность в выражении эмоций на снимках на этапе обучения нейронной сети для повышения качества распознавания. были сделаны уточнения к моделям, описанным в [6]. В дальнейшем отбор изображений и их разметка для обучения сети проводились в соответствии с приведенными ниже описаниями выражения эмоций.

Удивление: брови округлены и высоко подняты; горизонтальные морщины пересекают весь лоб; верхние веки подняты, а нижние – опущены; напряжения или натяжения в области рта нет, челюсть опущена так, что рот открыт, губы и зубы разъединены.

Счастье: уголки губ оттянуты в стороны и приподняты; рот может быть приоткрыт или закрыт, а зубы обнажены или прикрыты губами; от носа к внешнему краю губ (позади уголков губ) тянутся морщины – носогубные складки; щеки подняты; под нижними веками образуются морщинки, нижние веки могут быть приподняты, но не напряжены; у наружного края уголков глаз появляются морщинки («гусиные лапки»).

Гнев: брови опущены и сведены; между бровями появляются вертикальные складки; нижние веки напряжены и могут быть как подняты, так и оставаться без изменений; верхние веки напряжены и могут не изменить своего положения или же опуститься вследствие движения бровей; глаза неподвижны и могут казаться выпученными; губы находятся в одном из двух положений: а) крепко сжаты, причем уголки губ могут быть прямыми или опущенными или же б) раскрыты и напряжены в форме квадрата, как при крике; ноздри могут быть расширены.

Печаль: изогнутые вниз губы и поднимающиеся внутрь брови с образованием морщин на лбу; верхние веки слегка опущены; рассеянный взгляд, уголки рта слегка опущены; внутренние уголки бровей подняты вверх; кожа под бровями становится треугольной формы; внутренние уголки верхних век приподняты; уголки губ опущены или же губы расслаблены.

Изображения были предварительно масштабированы и обрезаны, а также переведены в оттенки серого. Для увеличения объема и разнообразия данных была проведена аугментация, в частности, к имеющимся снимкам был добавлен гауссовский шум в разных процентных отношениях к исходному изображению. Таким образом удалось увеличить датасет до 10000 изображений. Далее данные были разделены на выборки – обучающую, валидационную и тестовую в соотношении 70-15-15 %.

РЕЗУЛЬТАТЫ И ВЫВОДЫ

В результате проведенной работы установлено, что уточнение модели приводит к повышению числа правильно распознанных эмоций из тестовой выборки на 5-12 % в зависимости от вида эмоции. При этом на обучающей выборке результат достигал, как и в случае, описанном авторами [2], порядка 98%, однако на тестовой и валидационной – не превышал 79 и 64 % соответственно. Этот результат нельзя назвать отличным, однако он указывает на возможность повышения качества распознавания эмоций на основе комплексного подхода – уточнения модели эмоций, изменения архитектуры сети и расширения датасета методами аугментации и другими. В дальнейшем планируется расширить набор данных для обучения сети, применить различные архитектуры, а также функции активации, провести тестирование её работы на изображениях с несколькими лицами, а также с кадрами из видеопотока.

БИБЛИОГРАФИЧЕСКИЕ ССЫЛКИ

1. Бергер Е. Г., Бублик И. Ю. Анализ применения искусственного интеллекта на базе технологии отслеживания эмоций человека // *Colloquium-journal*. 2020. №10 (62).
2. Тимофеева Ольга Павловна, Неимущев Сергей Андреевич, Неимущева Любовь Ивановна, Тихонов Иван Александрович. Распознавание эмоций по изображению лица на основе глубоких нейронных сетей // *Труды НГТУ им. Р. Е. Алексеева*. 2020. №1 (128).
3. Рюмина Е.В., Карпов А.А. Аналитический обзор методов распознавания эмоций по выражениям лица человека // *Научно-технический вестник информационных технологий, механики и оптики*. 2020. №2.
4. Нейросетевой алгоритм распознавания эмоций человека по изображению лица [Электронный ресурс]. URL: <https://elib.bsu.by/bitstream/123456789/248683/1/245-248.pdf> (дата обращения 30.03.2024)
5. Сверточная нейронная сеть, часть 1: структура, топология, функции активации и обучающее множество [Электронный ресурс]. URL: <https://habr.com/ru/articles/348000/> (дата обращения 18.02.2024)
6. Экман П. Психология эмоций. Я знаю, что ты чувствуешь//изд. Питер; СПб. 2010. С. 235. ISBN 5–49807–705–5
7. Ali Mollahosseini, Behzad Hasani and Mohammad H. Mahoor. "AffectNet: A Database for Facial Expression, Valence, and Arousal Computing in the Wild," in *IEEE Transactions on Affective Computing*, vol. 10, no. 1, 1 January-March 2019. P. 18-31. DOI: 10.1109/TAFFC.2017.2740923.