

Определение эмоционального состояния обучаемого-ребенка в цифровой образовательной среде по статическим изображениям с применением нейросети

М.С. Дьяченко¹, А.Г. Леонов², М.В. Райко³, А.А. Холькина⁴

¹ФГУ ФНЦ НИИСИ РАН, Москва, Россия, mdyachenko@niisi.ru;

²ФГУ ФНЦ НИИСИ РАН, Москва, Россия, МГУ им. М. В. Ломоносова, Москва, Россия, МПГУ, Москва, Россия, Государственный университет управления, Москва, Россия, dr.l@vip.niisi.ru;

³ФГУ ФНЦ НИИСИ РАН, Москва, Россия, rayko@niisi.ru;

⁴ФГУ ФНЦ НИИСИ РАН, Москва, Россия, kholkina@niisi.ru

Аннотация. В статье описаны исследования возможности распознавания детских эмоций на основе сверточных нейронных сетей для использования в цифровых образовательных средах таких как, например, ЦОС “ПиктоМир”. Рассмотрены альтернативные решения и изложен опыт создания собственной нейронной сети, обученной на экспериментальном наборе данных. Приводятся результаты тестирования системы с помощью различных метрик на основе валидационного набора данных и сравнение полученных результатов с наиболее популярными алгоритмами распознавания эмоций. Сделаны выводы о направлении дальнейших исследований, в частности, необходимость расширения обучающего набора данных, предложены дальнейшие шаги к доработке с целью интеграции в образовательную среду “ПиктоМир”.

Ключевые слова: нейросети, глубокое обучение, распознавание эмоций, классификация эмоций, ПиктоМир.

1. Введение

В последнее время наблюдается интерес к компьютерным программам и системам, учитывающим эмоциональное состояние человека. Сфера применения компьютерного зрения для распознавания эмоций расширяется, и задача становится все более актуальной. На сегодняшний день системы распознавания находят применение в маркетинге, в системах безопасности, в мониторинге состояния водителей, в образовании, в банковской сфере и других областях [1, 2]. В рамках выполненного исследования предпринята попытка сконструировать систему распознавания эмоций детей на основе нейронной сети для использования в цифровой образовательной среде «ПиктоМир», функционирующей на массовых современных планшетах. Мы считаем, что автоматизированные цифровые наблюдения за мимическими эмоциональными реакциями учащегося позволят уловить его настроение, понять, когда он доволен, своими успехами (успешно решил задачу), расстроен из-за ошибок и т.п. Мы надеемся, что полученные данные об эмоциональном состоянии ребенка можно будет использовать для улучшения качества и эффективности обучения.

Авторы ставили целью дифференцировать различные эмоциональные состояния детей,

чтобы педагог смог вовремя определить, что ребенок устал, испытывает трудности при выполнении задания. Более того, на изменение настроения ученика может реагировать не только педагог (осведомленный средой ПиктоМир о необходимости вмешательства), но и сама среда ПиктоМир, рекомендую ребенку либо **выполнить** определенные действия вне работы в среде ПиктоМир (взять перерыв на отдых; попросить помощи у преподавателя и т.д.), либо выполнить автоматизированные действия в самой цифровой образовательной среде. Например, посмотреть мультфильм с подсказкой, пообщаться с виртуальным ассистентом и пр.

2. Обзор значимых публикаций и поиск известных решений

Для решения описанной выше задачи распознавания эмоций, часто используют нейросети – аналогичные тем, что используются для оперирования с аудио, видео и статическими изображениями. Согласно современным исследованиям высокую эффективность в подобных задачах имеют сверточные нейронные сети. Так существуют разработки, способные определять эмоциональное состояние человека по голосу. Например, одно из исследований, посвященное распознаванию эмоций по голосу [3], показало

точность сверточной нейросети 73%. Галичий, Афанасьев и Нестеров в своей работе [4] рассмотрели применение сверточных нейронных сетей для распознавания по статическим изображениям - фото. Нужно отметить, что исследования показали высокие результаты у простейших сверточных сетей, сравнимый со сложными дообученными моделями. К основным выявленным проблемам относятся дисбаланс классов, который необходимо учесть при создании собственного датасета, и переобучение. В аналогичном исследовании по распознаванию эмоций по фотографиям, проведенном на основе сверточной нейросети, точность составила 67,7% [5]. В этой статье указываются также такие проблемы, как отсутствие накопленной объемной базы данных изображений для обучения и трудности при переходе к распознаванию видео (не статичных) изображений.

В классе уже существующих интегрированных разработок был рассмотрен онлайн-сервис распознавания по видео “Emojify” [6], разработанный компанией “Dovetail Labs”. Нейросетевой алгоритм, заложенный в основу системы, позволяет распознавать в реальном времени 6 эмоций с точностью до 99%, используя камеру с телефона или компьютера. Отсутствие полноценной версии в свободном доступе ограничивает использование указанного решения. Аналогичным инструментарием является система, разработанная при поддержке МГУ им. Ломоносова, которая носит название “EmoDetect” [7]. Эта система позволяет распознавать те же 6 эмоций. Однако эта система распространяется в свободном доступе в виде бинарных исполняемых файлов только для ОС Windows, что является потенциальной проблемой для внедрения в “ПиктоМир” и в подобные цифровые образовательные среды.

В процессе поиска уже обученных нейросетей распознавания эмоций был найден набор инструментов “EmoPy” [8] с двумя нейросетями, которые классифицируют эмоции на основе изображений лиц людей. Обучение проводилось на датасете FER+ [9]. Точность нейросети “ConvolutionalLstmNN” на валидации составила 47%, у “TransferLearningNN” всего 29%.

Стоит отметить, что все рассмотренные системы относятся к распознаванию эмоций у взрослых. Эксперименты с распознаванием детских эмоций не описывались, поэтому на данном этапе нет возможности судить о применимости подобных решений относительно детских эмоций без потери качества распознавания.

3. Постановка задач исследования

Поскольку поиски готовых известных решений не привели к желаемому результату в виде готовой модели, которая бы распознавала детские эмоции с хорошей точностью и могла бы быть встроена в цифровую образовательную среду “ПиктоМир”, стала актуальной отдельная работа по разработке подобной модели. В ходе исследования было поставлено три задачи:

1. сформировать набор данных;
2. создать и обучить сверточную нейронную сеть;
3. оценить качество распознавания.

3.1. Набор данных

Для обучения нейросети был собран экспериментальный датасет с фотографиями одного ребенка, который регулярно занимается в образовательной среде “ПиктоМир”. Было получено письменное согласие родителей на проведение исследования и согласие на публикацию в научных статьях фотографий ребенка, сделанных в этом исследовании. Ребенок фотографировал себя на фронтальную камеру и располагал планшет так же, как на занятиях, выражая восемь видов эмоций. Съемка проходила в школе при искусственном и естественном освещении. Экспериментальный датасет включил в себя около 1600 фотографий и 200-230 фотографий на каждый из 8 классов эмоций. Список используемых классов эмоций включал:

1. удивление (surprise);
2. отвращение (disgust);
3. грусть (sad);
4. злость (angry);
5. страх (fear);
6. счастье (happy);
7. презрение (contempt);
8. нейтральное состояние (neutral).

Датасет был поделен на две части: обучающий набор (75%) и валидационный набор (25%). На первом подбирались веса модели, на втором были посчитаны метрики качества.

3.2. Выбор архитектуры нейросети и процесс обучения

Для реализации был выбран язык программирования Python и фреймворк машинного обучения Pytorch. Для решения задачи распознавания эмоций была выбрана предобученная модель на архитектуре сверточной нейронной сети ResNet152. Согласно исследованию [4], метод дообучения заранее обученных моделей сверточных нейросетей с большим количеством слоев (Transfer Learning) весьма эффективен. Модель была предварительно обучена на наборе

данных ImageNet. Он включает в себя более миллиона изображений и 1000 классов, одним из которых является класс «Человек». Она уже содержит веса и смещения, которые представляют особенности того набора данных, на котором она была обучена. Такая модель будет быстрее обучаться на новых данных.

Архитектура сети содержит 152 слоя (см.

Рис. 1.). Входной слой, на который подается изображение, имеет размер $224 \times 224 \times 3$, где первые два числа – это ширина и высота изображения, а третье – количество каналов. Далее идут слои со свертками, батч-нормализацией, функцией активации ReLU и макс-пуллингом.

слой	выход	152 слоя
свёртка1	112×112	$7 \times 7, 64$, шаг 2
свёртка1_x	56×56	3×3 макс-пул, шаг 2
		$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
свёртка2_x	28×28	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
		$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
свёртка3_x	14×14	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
		$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
свёртка4_x	7×7	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	средний пул, 1000-ПС

Рис. 1. Архитектура нейросети ResNet152

Выходной полносвязный слой в предобученной модели имеет 1000 классов, это значение было заменено на 8 классов нового датасета. Чтобы определить принадлежность изображения к классу, значения выходного слоя проходят через функцию аргмакс. Соответственно, результирующим классом является тот, который имеет наибольшее значение на выходе.

Для расширения датасета был применен метод аугментации. Этот метод позволяет добавить на изображения различные искажения, изменения контраста, яркости и сатурации, добавить шумы, повернуть изображения, случайным образом удалить маленькие куски фото. В данной

нейросети используется 18 функций аугментации.

Процесс тренировки модели длился 15 эпох с батчами по 32 изображения. В качестве оптимизатора был выбран оптимизатор Адам с гиперпараметром $\text{learning rate} = 0.0005$ (остальные по умолчанию) [10], в качестве функции потерь взята кросс-энтропия. Значение функции потерь на обучающем наборе равнялось 0.013 (см. рис. 2), на валидационном наборе – 0.045 (см. рис. 3). Ниже представлены графики поведения функции ошибки при обучении и валидации.

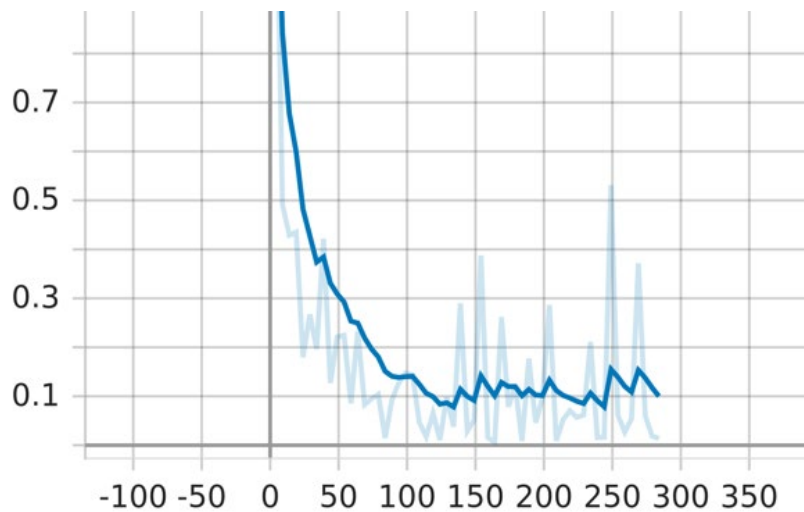


Рис. 2. График ошибки на обучающем наборе

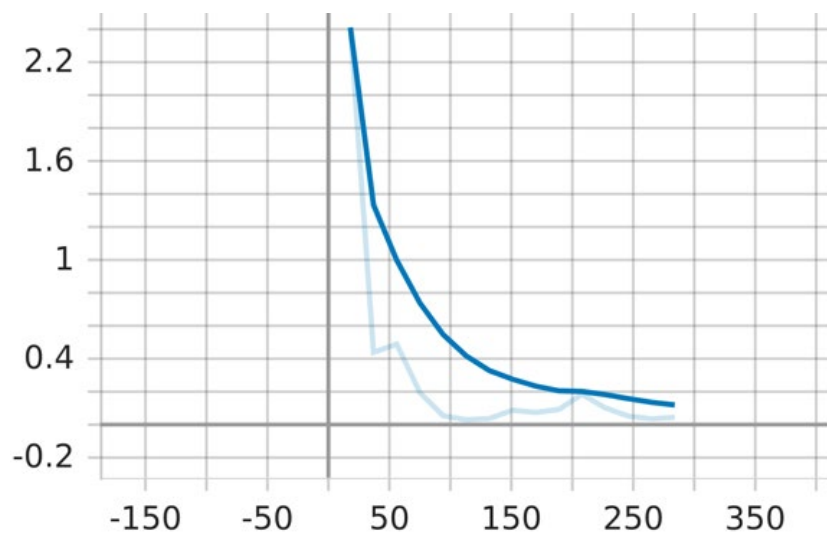


Рис. 3. График ошибки на валидационном наборе

По графикам видно, что у данной нейросети есть потенциал роста, и есть смысл продолжить тренировку для достижения более точных результатов.

3.3. Оценка качества распознавания

Для оценки качества распознавания использовались confusion matrix (матрица ошибок), precision (точность) и recall (полнота).

Для обученной модели распознавания эмоций была получена следующая матрица ошибок:

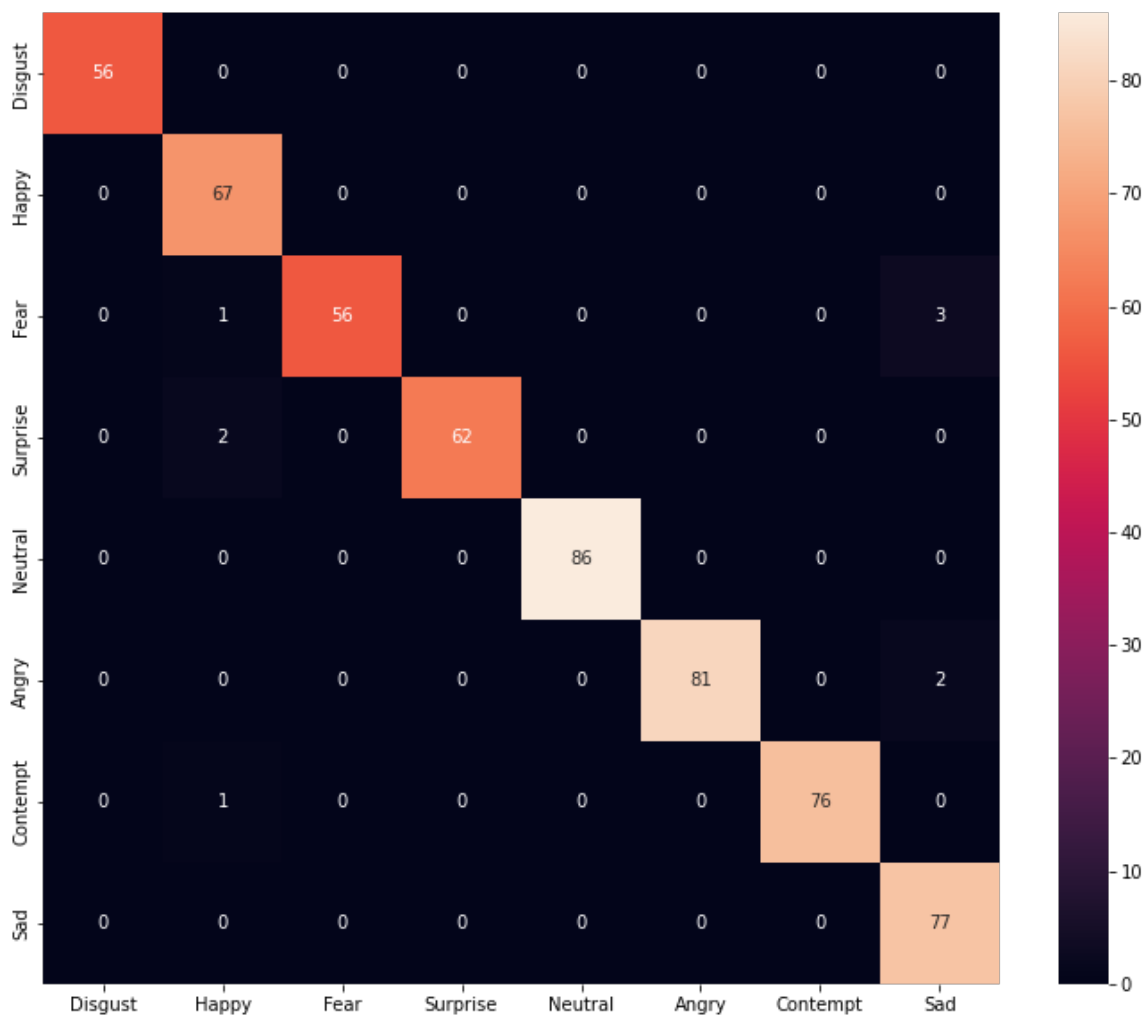


Рис. 4. Матрица ошибок

Напомним, что по горизонтали — прогноз модели, по вертикали — истинные классы. Можно заметить, что модель предсказывает довольно точно. Чаще всего модель ошибалась в

предсказании класса “Fear” - 4 раза, чуть меньше в классах “Surprise” и “Angry” - 2 раза, и один раз в классе “Contempt”. Ниже приведены примеры верной классификации.



Рис. 5. Предсказанный класс – “Happy”



Рис. 6. Предсказанный класс – “Neutral”



Рис. 7. Предсказанный класс – “Sad”

Для вычисления precision и recall были взяты средние значения по всем классам.

Mean precision = 0.985

Mean recall = 0.983

Значения говорят о высокой точности модели, но стоит отметить, что обучение проходило на небольшом объеме данных, поэтому при расширении обучающего датасета ожидается снижение точности. По итогам обучения можно сделать вывод, что архитектура нейросети ResNet152 подходит под задачу и её можно использовать в дальнейших исследованиях.

3. Результаты и выводы

В проведенном исследовании была проведена опытная практическая работа по созданию системы распознавания эмоций на основе сверточной нейронной сети. Выбранная модель зарекомендовала себя достаточно быстрым и эффективным дообучением. На валидационной выборке точность составила 98,5%, однако, стоит заметить, что обучение и проверка результатов

проводились на относительно небольшом датасете. В дальнейшем планируется расширение датасета, которое будет учитывать такие нюансы, как различное освещение, различные движения головы и различия в чертах лица из-за этнической принадлежности, возраста, пола, и очков, а также обучение нейросетей на других архитектурах для достижения еще большей точности распознавания.

Из-за ограниченности бюджета данного исследования, использовался датасет фотографий одного единственного ребенка. Безусловно, следует проверить результат работы построенной нейронной сети распознавания эмоций на достаточно представительном множестве детей возраста 6-8 лет. При достижении качественного распознавания (с точностью 70-80%), нейросеть планируется встроить в цифровую образовательную среду “ПиктоМир”.

Авторы благодарны администрации школы 199 г. Москвы и ученице 3 класса этой школы Арине Пушиной за помощь в организации и проведении исследования, описанного в настоящей статье.

Публикация выполнена в рамках государственного задания ФГУ ФНЦ НИИСИ РАН по теме № FNEF-2022-0010.

Determining the Emotional State of a Student-Child in a Digital Educational Environment From Static Images Using a Neural Network

M.S. Diachenko, A.G. Leonov, M.V. Rayko, A.A. Kholkina

Abstract. The article describes research on the possibility of recognizing children's emotions based on convolutional neural networks for use in digital educational environments such as, PiktoMir. Alternative solutions are considered and the experience of creating your own neural network trained on an experimental data set is presented. The results of testing the system using various metrics based on the validation data set and comparing the results with the most popular emotion recognition algorithms are presented. Conclusions are drawn on the direction of further research, in particular, the need to expand the training data set, further steps are proposed for refinement in order to integrate into the PiktoMir educational environment.

Keywords: neural networks, deep learning, emotion recognition, classification of emotions, PiktoMir.

Литература

1. Е. Хрисанфова. Независимое издание о технологиях и бизнесе RB.RU [Электронный ресурс]: Эмоциональный ИИ: кто и зачем распознаёт эмоции в России и за рубежом. <https://rb.ru/longread/emotion-ai/>.
2. Ю. Н. Харари. 21 урок для XXI века. Синдбад, 2019. с. 52.
3. А. Сергунов, Д.И., А.А. Артёмова и С.С. Гришунов. Система распознавания эмоций по голосу на основе сверточной нейронной сети. E-Scio, 7, 2019. <https://cyberleninka.ru/article/n/sistema-raspoznaniya-emotsiy-po-golosu-na-osnove-svertochnoy-neyronnoy-seti/viewer>.
4. Галичий, Д.А., Г.И. Афанасьев и Ю.Г. Нестеров. Распознавание эмоций человека при помощи современных методов глубокого обучения. E-Scio, 5, 2019. <https://cyberleninka.ru/article/n/raspoznvanie-emotsiy-cheloveka-pri-pomoschi-sovremennyh-metodov-glubokogo-obucheniya/viewer>.
5. M. Singh, S. K. Sharma, S. Paul, J. P. Sajeevan, S. Paul. Facial emotion recognition system. IJARIE-ISSN(O)-2395-4396, Vol-6 Issue-6, 2020. https://www.researchgate.net/publication/347797356_FACIAL_EMOTION_RECOGNITION_SYSTEM
6. Сайт сервиса “Емоjify” [Электронный ресурс]. <https://emojify.info/menu>. (дата обращения: 01.08.2022)
7. Сайт сервиса “ЕмоDetect” [Электронный ресурс]. <https://emodetect.ru>. (дата обращения: 01.08.2022)
8. Репозиторий “ЕмоPy” [Электронный ресурс]. <https://github.com/thoughtworksarts/ЕмоPy>. (дата обращения: 01.08.2022)
9. Набор данных с эмоциями “FER+” [Электронный ресурс]. <https://github.com/Microsoft/FER-Plus>. (дата обращения: 01.08.2022)
10. Оптимизатор Adam [Электронный ресурс]. <https://pytorch.org/docs/stable/generated/torch.optim.Adam.html>. (дата обращения: 01.08.2022)
11. Иванов, И.А., Е.А. Сопов, И.А. Панфилов. Многокритериальный подход к проектированию ансамбля нейросетевых классификаторов с отбором информативных признаков для решения задачи распознавания эмоций. Сибирский аэрокосмический журнал, 4, 2015. <https://cyberleninka.ru/article/n/mnogokriterialnyy-podhod-k-proektirovaniyu-ansamblya-neyrosetevyh-klassifikatorov-s-otborom-informativnyh-priznakov-dlya-resheniya/viewer>.
12. Курицкий, В.Ю., С.В. Садов. Нейросетевой алгоритм распознавания эмоций человека по изображению лица. Компьютерные технологии и анализ данных (СТДА'2020): материалы II Междунар. науч.-практ. конф., Минск, 2020: БГУ <https://elib.bsu.by/bitstream/123456789/248683/1/245-248.pdf>.