

Поиск редких данных в задаче распознавания лиц на изображениях

А.Д. Соколова¹, А.В. Савченко¹, С.И. Николенко²

¹Национальный исследовательский университет Высшая школа экономики,
Лаборатория алгоритмов и технологий анализа сетевых структур,
603093, Россия, г. Нижний Новгород, ул. Родионова, д. 136,

²Санкт-Петербургский государственный университет,
199034, Россия, г. Санкт-Петербург, Университетская наб., д. 7-9

Аннотация

Одной из основных проблем современных нейросетевых дескрипторов в задаче идентификации лиц является малое число обучающих примеров определенного типа: изображения плохого качества, разный масштаб или освещение, лица детей, пожилых людей, редкие расы. В результате точность распознавания оказывается низкой для входных изображений, не похожих на большинство изображений в наборе данных, используемом для настройки метода извлечения признаков. В работе предлагается способ преодоления такой проблемы за счет автоматического обнаружения нетипичных входных изображений на основе введения предварительного этапа их автоматической отбраковки. Для этого используется специальная свёрточная сеть, обученная на наборе редких данных, которые обрабатывались с помощью известных алгоритмов преобразования изображений. Для повышения вычислительной эффективности решение о наличии редкого изображения принимается на основе того же дескриптора лица, который используется в классификаторе. Экспериментальное исследование подтвердило преимущества в точности предложенного подхода для нескольких наборов данных лиц и современных нейросетевых дескрипторов.

Ключевые слова: распознавание лиц, обнаружение аномалий, обработка изображений, распознавание редких изображений.

Цитирование: Соколова, А.Д. Поиск редких данных в задаче распознавания лиц на изображениях / А.Д. Соколова, А.В. Савченко, С.И. Николенко // Компьютерная оптика. – 2022. – Т. 46, № 5. – С. 801–807. – DOI: 10.18287/2412-6179-CO-1061.

Citation: Sokolova AD, Savchenko AV, Nikolenko SI. Open-set face identification with automatic detection of out-of-distribution images. Computer Optics 2022; 46(5): 801-807. DOI: 10.18287/2412-6179-CO-1061.

Введение

Распознавание и идентификация лиц на протяжении долгого времени является одной из наиболее актуальных и сложных задач компьютерного зрения. В то же время всё ещё остаётся большое число нерешённых проблем, связанных с низким качеством входных данных: масштаб, яркость, разрешение и т. д. Кроме того, внешность человека со временем меняется: появляются признаки возрастных изменений, изменение причёски, макияжа, оттенка кожи и другое. Благодаря появлению дескрипторов [1], основанных на глубоких свёрточных нейронных сетях (СНС), и наличию больших наборов данных для обучения [2] удалось достичь значительного повышения качества распознавания даже в неконтролируемых варьирующихся условиях наблюдения. К сожалению, поскольку нейросетевые дескрипторы обычно обучаются на больших наборах лиц знаменитостей [3], алгоритмы их классификации допускают ошибки при принятии решения на изображениях, не типичных для подобных выборок (проблема «длинных хвостов» распределений, от англ. “long tail”). В результате существующие алгоритмы допускают ошибки при распознавании редких (англ. “out-of-distribution”) входных изображений, значительно отличающихся от большинства

изображений в наборе данных, используемом для обучения алгоритма извлечения признаков [4].

В настоящей статье предлагается способ преодолеть отмеченную проблему за счет разработки детектора аномальных изображений лиц, который используется на предварительном этапе обработки. Если задача состоит в распознавании статических изображений, то решение для обнаруженных детектором данных не принимается. При покадровой обработке видеопотока аномальным кадрам может присвоиться более низкий вес для принятия окончательного решения [5]. Предложен автоматический способ сбора специального набора данных из множества фотографий, используемого для обучения дескрипторов лиц, который используется для обучения нейронной сети, определяющей аномальные изображения. Приведённые результаты и сделанные по ним выводы представляют интерес для широкого круга специалистов в области распознавания образов.

Постановка задачи

Задача распознавания лиц на открытом множестве (от англ. open-set) состоит в соотнесении входного изображения с одним из заранее определенных людей (классов) $C > 1$ или откладывании решения о человеке, изображений лица которого потенциально нет в обу-

чающей выборке [6]. Предполагается, что каждый класс (человек) задается набором его фотографий, то есть доступна обучающая выборка с $R \geq C$ примерами, причем для каждого r -го изображения известна метка класса $c(r) \in \{1, \dots, C\}$. Так как обучающий набор данных обычно содержит мало примеров для каждого из C классов (иногда одна фотография каждого человека) [4], в настоящее время обычно используются методы доменной адаптации [7]. В них на предварительном этапе сверхбольшой набор данных лиц знаменитостей применяется для обучения глубокой СНС, которая далее используется для извлечения характерных признаков лиц во входном изображении [2, 8]. Для этого изображение лица подается на вход СНС, а выходы её предпоследнего слоя формируют вектор признаков (дескриптор) $x = [x_1, \dots, x_D]$ высокой размерности $D > 1$. Зачастую их нормируют, используя L_2 норму, чтобы обеспечить дополнительную устойчивость к изменчивости условий, а также сравнивать дескрипторы в метрике Евклида вместо косинусной меры близости [9]. Далее каждое r -е изображение обучающего множества подается на вход той же СНС для извлечения его вектора признаков $x_r = [x_{r,1}, \dots, x_{r,D}]$. При наличии малого числа примеров каждого класса обычно применяют метод ближайших соседей (k-NN) [10].

Для извлечения характерных признаков лиц использованы следующие дескрипторы:

- 1) СНС ResNet50 (VGGFace-2) [2], извлекающая дескриптор размерности $D=2048$ со слоя «pool5_7x7_s1»;
- 2) СНС Multi-output MobileNetV1 [8], выдающая вектор признаков размерности $D=1024$ со слоя «reshape_1/Mean»;
- 3) СНС InsightFace (ArcFace) [9], в которой со слоя «fc1_output» получается $D=512$ -мерный дескриптор;
- 4) СНС FaceNet [11], извлекающая вектор размерности $D=512$ со слоя «embeddings».

Сложность задачи идентификации лиц определяется различным числом ракурсов и качества распознаваемых изображений лиц. Особое влияние на точность оказывают входные изображения, которые не характерны для большинства примеров из обучающего набора (рис. 1). Мы будем называть их «редкими» изображениями (англ. «out-of-distribution») [12], потому что они значительно отличаются от изображений знаменитостей, которые использовались для обучения СНС [2, 3]. Из-за большого разнообразия входных изображений одного и того же класса становится трудно правильно их классифицировать за счет слишком большой разницы с эталонными изображениями. Экспериментально было замечено, что дескрипторы таких изображений даже для различных людей, например, детей или пожилых людей различных рас [13], обычно близки по расстоянию друг к другу. Таким образом, основная задача настоящей

статьи состоит в том, чтобы автоматически проверить, является ли входное изображение редким/аномальным и может ли его распознавание привести к ухудшению итогового качества. Имея возможность определять «плохие» изображения, можно исключать их, тем самым улучшить качество классификации или понизить их влияние на итоговое решение при идентификации лиц на видео.

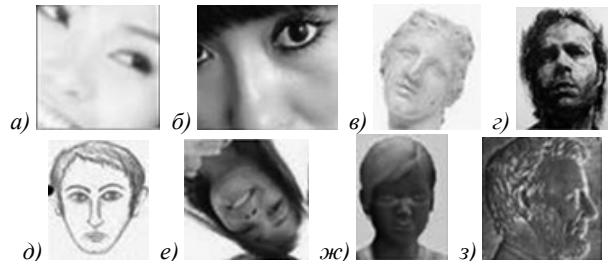


Рис. 1. Примеры редких изображений лиц

В задачах распознавания лиц ключевой показатель качества – точность, оцениваемая как доля корректно классифицированных лиц среди тестовой выборки. К сожалению, если предусмотрена возможность отказа от принятия решений (англ. reject option), точность может быть повышена за счёт исключения большей части изображений из тестового множества. В таком случае следует рассматривать многокритериальную постановку задачи, в которой дополнительно учитывается вероятность ложной тревоги (FPR, от англ. False Positive Rate) – доля ошибочно отклоненных (признанных «редкими») «нормальных» изображений. Кроме того, интерес представляет и оценка вероятности пропуска «редких» данных (FNR, от англ. False Negative Rate). Так как для разных прикладных задач может потребоваться проконтролировать различные метрики, в настоящей работе не делается попытка объединить их в показатель «итогового качества». В рамках экспериментального исследования оцениваются все указанные метрики, и рассматриваются Парето-оптимальные решения с акцентом на области с высокой итоговой точностью распознавания. При этом, изменяя параметры детектора «редких» фотографий (бинарного классификатора), можно понижать метрику FPR, повышая тем самым FNR и снижая итоговую точность распознавания.

Обзор алгоритмов нахождения редких данных

Для нахождения аномальных изображений могут применяться известные методы оценки их качества, такие как вычисление пикового отношения сигнала к шуму (PSNR) [14], определение яркости изображения и т.д. PSNR – это соотношение между максимально возможной мощностью сигнала и мощностью искажающего шума, которое помогает находить искажения изображения или различные артефакты. Другой показатель – структурное подобие (SSIM) [15], в котором измерение или прогноз качества изображения основывается на исходном несжатом или неискажен-

ном изображении в качестве эталона. Такие способы помогают идентифицировать только определенные искажения изображения, но было бы полезно знать, как обнаружить все аномалии.

Методы на основе автокодировщика [16] находят аномалии путем представления изображений в более низкую размерность, где моделируется фактическое распределение. Такой подход позволяет модели охватывать мультимодальные распределения во входном пространстве, но на практике модель дает размытые реконструкции. Еще один широко используемый подход для обнаружения аномалий – метод One-Class SVM [17], в котором машина опорных векторов изучает гиперплоскость для отделения всех точек данных от начала координат в гильбертовом пространстве воспроизводящего ядра и максимизирует расстояние от этой гиперплоскости до источника. На её основе была создана нейронная сеть One-Class [18], в которой признаки изображения особым образом переопределяются для аномальных изображений в сложных наборах данных, где граница между нормальными и аномальными областями очень нелинейна.

На данный момент времени существуют работы, демонстрирующие повышение точности моделей машинного обучения за счет удаления входных примеров, которые в большой степени влияют на качество основной модели [19]. Поскольку глубокие нейронные сети показывают высокие результаты в широком спектре приложений [20], они также используются для обнаружения редких данных. В одной из первых таких работ представлен простой подход, который использует вероятности из распределений softmax на выходе заданной нейронной сети [21]. В статье [22] оценивается степень уверенности (англ. confidence score) на основе расстояния Махаланобиса с помощью условных гауссовских распределений классов, при этом учитываются особенности глубоких моделей при гауссовском дискриминантном анализе. Недостатком таких подходов является менее практичный детектор, поскольку необходимо настраивать гиперпараметры, используя «редкие» выборки из тестового распределения. В статье [23] классификатор обучался с помощью генеративных состязательных сетей (GAN, от англ. Generative Adversarial Network) для создания редких образцов. Для каждого тестирования классификатор и GAN обучаются на сгенерированных редких данных: классификатор с большей точностью определяет редкие данные, а GAN генерирует эффективные обучающие образцы таких данных. Таким образом, модель может обнаруживать только определенные типы тестовых распределений аномалий.

Таким образом, все подобные работы определяют редкое изображение как новое входное изображение для модели, такое как синтезированный шум или специфичное распределение изображения. В настоящей же работе основной задачей является определение не только плохого качества или наличия дефек-

тов (рис. 1), но также наличие изображений людей разного возраста, простые изменения внешнего вида, небольшие изменения качества изображения и т.д. К сожалению, существующие базы данных больше не подходят для такого широкого диапазона малораспространенных изображений лиц. Набор Labeled Faces in the Wild [24] или набор IARPA Janus Benchmark [25] обычно используются в качестве тестовых баз данных, но они не включают существенное расовое [26], возрастное или внешнее разнообразие.

Предложенный подход

Для составления большого набора данных с редкими изображениями использовались достаточно большие базы VGGFace2 [2] и MS-Celeb-1M [3]. Набор VGGFace2 содержит 3,31 миллиона изображений 9131 человека, в то время как набор MS-Celeb-1M состоит из 10 миллионов изображений 100 000 человек. Для фотографий каждого человека была выполнена кластеризация с помощью метода k-means на 50 и 20 кластеров для VGGFace2 и MS-Celeb-1M соответственно. Далее из каждого кластера для человека была наугад выбрана одна фотография и помещена в обучающее множество, а все остальные изображения попали в тестовое множество, которое классифицировалось с помощью метода ближайших соседей. Изображения, на которых классификатор ошибается чаще всего, были выбраны в качестве редких (“out-of-distribution”) изображений. Если точность определения изображений конкретного человека больше 95%, то изображения обозначались как «нормальные», соответствующие распределению входных данных, на которых был обучен дескриптор.

Далее использовались наборы данных с изображениями людей разного возраста All ages faces dataset [27] (13322 изображения 13322 человек возраста от 2 до 80 лет) и Large age gap face verification [28] (3828 фотографий 1010 человек текущего возраста и их в молодости). Изображения детей (до 14 лет) и пожилых людей (после 55 лет) из них обозначались как «редкие», так как большинство существующих наборов данных [2, 3], используемых в обучении современных алгоритмов распознавания [1], в основном содержат фотографии знаменитостей среднего возраста.

Наконец, для получения искаженных редких данных (рис. 2) к выделенным «нормальным» изображениям были применены дополнительные искусственные трансформации [29], реализованные с помощью библиотеки OpenCV, такие как наложение нормально распределенного шума; размытие изображения; преобразование изображения в оттенки серого; поворот изображения; обрезка изображения; преобразование изображения в рисунок карандашом; преобразование изображения в нарисованное красками изображение. Кроме того, применялись существующие нейросетевые технологии [30, 31]: перенос макияжа; портретная манипуляция (смена возраста). Данный набор

преобразований может быть расширен, тем самым будет больший обхват «редких» изображений.

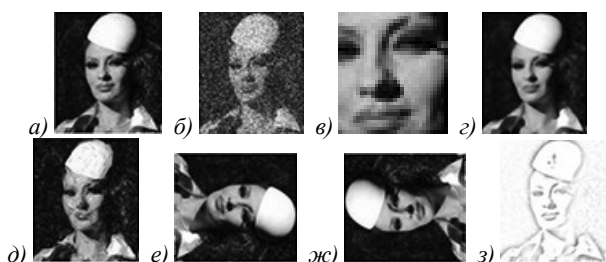


Рис. 2. Пример трансформаций

В результате, используя вышеупомянутые подходы, был собран сбалансированный набор данных из 150 000 «нормальных» и 150 000 «редких» изображений. С использованием этого набора исследовались несколько подходов для бинарной классификации «редких» и «нормальных» фотографий. Для этого применялись СНС с предварительно обученными на ImageNet весами, а также глубокая состязательная генеративная сеть (GAN), обученная так, чтобы генератор мог создавать реалистичные изображения (рис. 3), а дискриминатор мог быть использован для нахождения «редких» изображений. Кроме того, в настоящей работе предложено классифицировать тот же дескриптор лица, извлеченный СНС [2, 8, 9], который применяется в дальнейшем при распознавании лица. Помимо обычных классификаторов из библиотеки scikit-learn, была реализована простая полносвязная нейронная сеть. Её архитектура показана на рис. 4.



Рис. 3. Пример сгенерированных изображений

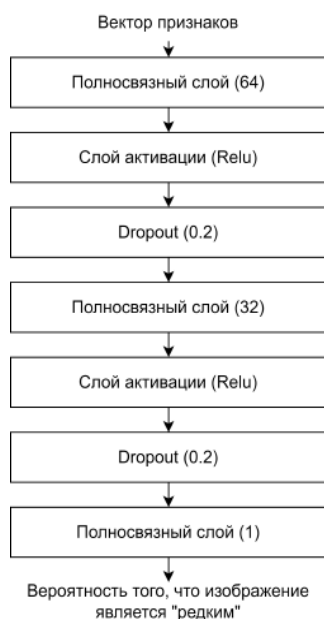


Рис. 4. Архитектура предложенного детектора редких кадров

Структурная схема итогового предлагаемого алгоритма представлена на рис. 5. Здесь в начале для поступившего на вход изображения детектируются области лиц с помощью модели MobileNet SSD из дескриптора лица, используя предварительно обученные СНС. На следующем этапе с помощью классификатора (рис. 4) проверяется, относится ли изображение к редким данным. Если изображение было признано «редким», решение не принимается (откладывается), в противном случае выполняется классификация лица с помощью метода ближайших соседей (k-NN).

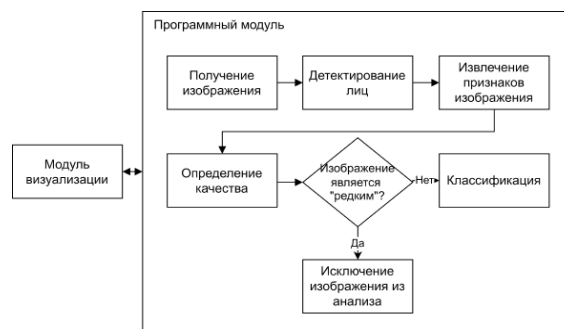


Рис. 5. Предложенный алгоритм

Экспериментальное исследование

Все эксперименты и обучение моделей проходили с использованием библиотек TensorFlow 2 и scikit-learn. Собранный набор «редких» и «нормальных» изображений был разбит пополам, первая половина использовалась для проведения сравнительного анализа различных детекторов аномалий, а вторая – для тестирования точности распознавания лиц.

В первый поднабор вошли 26 000 «нормальных» изображений и 23 000 трансформированных изображений (одинаковое число изображений на каждую трансформацию) из набора VGGFace2 [2], 50 000 «нормальных» изображений и 41 000 трансформированных изображений (одинаковое число изображений на каждую трансформацию) из набора MS-Celeb-1M [3], а также полные наборы данных All ages faces dataset [27] и Large age gap face verification [28]. Методы решения задачи бинарной классификации (детектирование «редких» изображений) обучались на 70 % этого поднабора, в то время как оставшиеся 30 % фотографий использовались для их валидации. Первый эксперимент заключался в использовании традиционного подхода с обучением бинарных классификаторов встроенных в TensorFlow 2 архитектуры EfficientNetB3, MobileNetV1, ResNet50, Xception, InceptionV3, предварительно обученных на изображениях из набора данных ImageNet. Обучение проходило в течение 20 эпох с ранним остановом с использованием оптимизатора Adam. К сожалению, такой подход не может быть использован в задаче детектирования «редких» изображений из-за переобучения: точность лучших моделей EfficientNet-B3 и ResNet-50 составила 58,9 % и 58,8 % соответственно, в то время

как для обучающих данных точность всех моделей составляет более 90 %.

Дополнительно были реализованы другие известные методы поиска аномалий. Генеративная модель GAN, обученная в течение 100 эпох, показала чуть более высокую, но все еще недостаточную для практики точность 69,4 %. Кроме того, использовался метод [19], в котором «редкие» экземпляры выводятся путем повторного отслеживания шагов стохастического градиентного спуска с включением промежуточных моделей, вычисляемых на каждом шаге. Обучение в течение 300 эпох позволило достичь точности 73,4 %.

Наилучшие результаты в этой задаче (табл. 1) были достигнуты с помощью применения алгоритмов классификации к известным дескрипторам лиц ResNet50 [2], MobileNetV1 [8], InsightFace [9], FaceNet [11]. Сравнивалось 5 методов классификации: k-NN, SVM (Support vector machine), Random Forest, LDA (линейный дискриминантный анализ, англ. Linear Discriminant Analysis) и полносвязная нейронная сеть (рис. 4). Лучшая точность на тестовых данных (94,8 %) достигается с использованием многозадачной модели MobileNetV1 [8] и предложенной полносвязной модели (рис. 4). Метод ближайшего соседа даёт сравнимую точность, однако является намного более сложным с вычислительной точки зрения за счёт необходимости хранения и перебора всех векторов признаков.

Табл. 1. Точность (%) детектора редких изображений для предобученных дескрипторов лиц

Классификатор	Res-Net50	Mobile-NetV1	Insight-Face	FaceNet
k-NN	89,7%	93,9%	89,6%	90,9%
SVM	69,1%	79,4%	63,0%	64,3%
Random Forest	85,0%	86,5%	80,6%	84,2%
LDA	88,1%	89,0%	83,7%	87,4%
Полносвязная нейронная сеть (рис. 4)	92,3%	94,8%	89,7%	90,2%

Показатели FPR и FNR, характеризующие вероятности ошибок 1 и 2 рода для детектора редких данных, приведены в табл. 2. Как видно из этих результатов, только около 2 % «нормальных» изображений были определены как «редкие».

Табл. 2. Оценки метрик FPR и FNR для детектора редких данных

Набор данных	СНС	FPR	FNR
VGGFace2	ResNet50	6,79%	2,71%
	MobileNetV1	1,86%	1,56%
	InsightFace	10,43%	3,45%
	FaceNet	9,48%	5,27%
MS-Celeb-1M	ResNet50	7,54%	3,46%
	MobileNetV1	8,43%	2,72%
	InsightFace	11,05%	6,84%
	FaceNet	9,31%	5,45%

В следующем эксперименте сравнивалась точность распознавания лиц на всех изображениях с

предлагаемым подходом, где классифицируются только «нормальные» изображения, выбранные с помощью полносвязной нейронной сети (рис. 4), а остальные фотографии *исключаются* из анализа. Кроме того, исследован гибридный подход, в котором дескрипторы лиц «редких» изображений *преобразовывались* следующим образом: к вектору признаков найденного «редкого» изображения прибавляется разность между средними дескрипторами «нормальных» и «редких» изображений для того, чтобы компенсировать различия в средних значениях дескрипторов двух разных классов.

Так как часть изображений использовалась для обучения детектора «редких» изображений, тестирование осуществлялось на второй половине изображений, взятых из набора данных VGGFace2, а именно 26 172 обычных и 23 653 изображения, которые получены с помощью трансформаций (рис. 3) $C = 188$ человек.

Результат классификации для наиболее точных методов представлен в табл. 3. Метод ближайших соседей с использованием MobileNetV1 [8] для извлечения признаков изображения лица демонстрирует наивысшую точность при исключении редких изображений. Такой подход оказался в 2–4 раза точнее по сравнению с классификатором Random Forest. Линейный дискриминантный анализ уступает алгоритму k-NN по точности около 1 % для разных дескрипторов. Преобразование «редких» изображений не привело к значительным улучшениям (около 0,5 % в зависимости от классификатора и дескриптора лиц).

Табл. 3. Точность (%) классификации изображений VGGFace2

Алгоритм	Res-Net50	Mobile-NetV1	Insight-Face	Face-Net
k-NN + все	88,7	94,9	83,7	94,1
k-NN + исключение	92,4	97,9	83,9	97,7
k-NN + преобразование	89,3	94,9	83,8	94,4
Random Forest + все	36,5	21,2	19,9	42,9
Random Forest + исключение	82,7	91,6	58,5	93,5
Random Forest + преобразование	37,3	21,8	19,7	46,1
LDA + все	88,6	92,6	75,7	92,5
LDA + исключение	93,3	97,1	82,0	96,8
LDA + преобразование	88,8	93,0	75,9	92,4

Для подмножества выбранного набора данных из MS-Celeb-1M, содержащего 50 597 обычных изображений $C = 320$ человек и их 41 665 трансформированных изображений, результаты приведены в табл. 4. Алгоритм k-NN демонстрирует самую высокую точность для данного набора данных при использовании дескриптора ResNet50, обученного на наборе данных VGGFace2 [2]. Точность линейного дискриминантного анализа оказалась ниже точности k-NN.

Табл. 4. Точность (%) классификации изображений MS-Celeb-1M

Алгоритм	Res-Net50	Mobile-NetV1	Insight-Face	Face Net
k -NN + все	87,6	86,3	83,8	84,2
k -NN + исключение	91,9	90,5	84,1	89,3
k -NN + преобразование	87,7	86,3	84,0	84,4
Random Forest + все	50,1	37,4	43,2	32,3
Random Forest + исключение	79,8	85,7	66,5	79,9
Random Forest + преобразование	50,1	37,2	41,7	32,3
LDA + все	88,7	86,4	67,7	79,8
LDA + исключение	91,5	91,0	67,9	82,3
LDA + преобразование	88,6	86,6	67,5	79,7

Заключение

В настоящей статье был предложен эффективный алгоритм распознавания лиц (рис. 5), в котором нейросетевой дескриптор лиц используется не только для принятия окончательного решения в методе ближайшего соседа, но и для обнаружения редких входных изображений с помощью полносвязной нейронной сети (рис. 4). Экспериментальные исследования продемонстрировали, что использование метода ближайших соседей с исключением изображений, обнаруженных детектором редких данных, показывает самые высокие показатели точности для дескрипторов лиц ResNet50 [2] и MobileNetV1 [8]. При этом извлечение признаков с помощью последней СНС обычно значительно быстрее, поэтому она является более подходящим способом извлечения признаков для практического применения в режиме реального времени, например, при распознавании лиц на видео.

В будущем необходимо разработать более подходящий способ применения информации о редких данных без их отбраковки, например, для задач обработки видеоданных. Кроме того, следует исследовать применение предложенного подхода для известной проблемы расовой предвзятости традиционных нейросетевых дескрипторов [26]: расы, слабо представленные в обучающем наборе фотографий знаменитостей, оказываются плохо различимы.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда (проект No 20-71-10010). Исследование Николенко С.И. поддержано Санкт-Петербургским государственным университетом, проект № 73555239 «Искусственный интеллект и наука о данных: теория, технология, отраслевые и междисциплинарные исследования и приложения».

References

[1] Deng J, Guo J, Xue N, Zafeiriou S. ArcFace: Additive angular margin loss for deep face. 2019 IEEE/CVF Conf on Computer Vision and Pattern Recognition (CVPR) 2019; 4690-4699.

- [2] Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A. Vgg-face2: A dataset for recognizing faces across pose and age. 13th IEEE Int Conf on Automatic Face & Gesture Recognition (FG 2018) 2018; 67-74.
- [3] Guo Y, Zhang L, Hu Y, He X, Gao J. MS-Celeb-1M: A dataset and benchmark for large-scale face recognition. In Book: Leibe B, Matas J, Sebe N, Welling M, eds. Computer Vision – ECCV 2016. Cham: Springer; 2016: 87-102.
- [4] Lahasan B, Lutfi SL, San-Segundo R. A survey on techniques to handle face recognition challenges: occlusion, single sample per subject and expression. Artif Intell Rev 2019; 52(2): 949-979.
- [5] Nikitin MYu, Konushin VS, Konushin AS. Neural network model for video-based face recognition with frames quality assessment. Computer Optics 2017, 41(5): 732-742.
- [6] Gunther M, Cruz S, Rudd EM, Boulton TE. Toward open-set face recognition. 2017 IEEE Conf on Computer Vision and Pattern Recognition Workshops (CVPRW) 2017: 573-582.
- [7] Xie H, Du Y, Yu H, Chang Y, Xu Z, Tang Y. Open set face recognition with deep transfer learning and extreme value statistics. Int J Wavelets Multiresolut Inf Process 2018; 16(4): 1850034.
- [8] Savchenko AV. Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet. PeerJ Comput Sci 2019; 5: e197.
- [9] Sokolova AD, Savchenko AV. Computation-efficient face recognition algorithm using a sequential analysis of high dimensional neural-net features. Optical Memory and Neural Networks 2020; 29(1): 19-29.
- [10] Faizov BV, Shakhuro VI, Sanzharov VV, Konushin AS. Classification of rare road signs. Computer Optics 2020, 44(2): 236-243.
- [11] Savchenko AV, Belova NS. Unconstrained face identification using maximum likelihood of distances between deep off-the-shelf features. Expert Syst Appl 2018; 108: 170-182.
- [12] Yu C, Zhu X, Lei Z, Li SZ. Out-of-distribution detection for reliable face recognition. IEEE Signal Process Lett 2020; 27: 710-714.
- [13] Zhou E, Cao Z, Yin Q. Naive-deep face recognition: Touching the limit of LFW benchmark or not? arXiv Preprint. Source: <https://arxiv.org/abs/1501.04690>.
- [14] Gupta D, Ahmad M. An efficient method to get improved peak signal to noise ratio (PSNR), using support vector machine. Int J Emerg Technol Adv Eng 2017; 7(9): 49-53.
- [15] Hore A, Ziou D. Image quality metrics: PSNR vs. SSIM. Proc 20th Int Conf on Pattern Recognition (ICPR) 2010: 2366-2369.
- [16] Rezende DJ, Mohamed S, Wierstra D. Stochastic back-propagation and approximate inference in deep generative models. Proc 31st Int Conf on Machine Learning (PMLR) 2014: 1278-1286.
- [17] Li KL, Huang HK, Tian SF, Xu W. Improving one-class SVM for anomaly detection. Proc Int Conf on Machine Learning and Cybernetics 2003; 5: 3077-3081.
- [18] Chalapathy R, Menon AK, Chawla S. Anomaly detection using one-class neural networks. arXiv Preprint. Source: <https://arxiv.org/abs/1802.06360>.
- [19] Hara S, Nitanda A, Maehara T. Data cleansing for models trained with sgd. arXiv Preprint. Source: <https://arxiv.org/abs/1906.08473>.
- [20] Sokolova AD, Savchenko AV. Organizing data in video surveillance systems based on deep learning technologies. In Book: Proceedings of the IV International Conference "Information Technologies and Nanotechnolo-

- gies" (ITNT 2018). Samara: "Novaja Tehnika" Publisher; 2018: 946-952.
- [21] Hendrycks D, Gimpel K. A baseline for detecting misclassified and out-of-distribution examples in neural networks. arXiv Preprint. Source: <https://arxiv.org/abs/1610.02136>.
- [22] Lee K, Lee K, Lee H, Shin J. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. arXiv Preprint. Source: <https://arxiv.org/abs/1807.03888>.
- [23] Lee K, Lee K, Lee H, Shin J. Training confidence-calibrated classifiers for detecting out-of-distribution samples. arXiv Preprint. Source: <https://arxiv.org/abs/1711.09325>.
- [24] Huang GB, Mattar M, Berg T, Learned-Miller E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Workshop on faces in 'Real-Life' Images: detection, alignment, and recognition, 2008: 1-14. Source: <https://hal.inria.fr/inria-00321923/document>.
- [25] Klare BF, Klein B, Taborsky E, Blanton A, Cheney J, Allen K, Jain AK. Pushing the frontiers of unconstrained face detection and recognition: IARPA Janus Benchmark A. 2015 IEEE Conf on Computer Vision and Pattern Recognition (CVPR) 2015: 1931-1939.
- [26] Wang M, Deng W, Hu J, Tao X, Huang Y. Racial faces in the wild: Reducing racial bias by information maximization adaptation network. 2019 IEEE/CVF Int Conf on Computer Vision (ICCV) 2019: 692-702.
- [27] Cheng J, Li Y, Wang J, Yu L, Wang S. Exploiting effective facial patches for robust gender recognition. Tsinghua Sci Technol 2019; 24(3): 333-345.
- [28] Bianco S. Large age-gap face verification by feature injection in deep networks. Pattern Recognit Lett 2017; 90: 36-42.
- [29] Nikolenko SI. Synthetic data for deep learning. Cham: Springer Nature Switzerland AG; 2021.
- [30] Chen YC, Lin H, Shu M, Li R, Tao X, Shen X, Jia J. Facelet-bank for fast portrait manipulation. 2018 IEEE/CVF Conf on Computer Vision and Pattern Recognition 2018: 3541-3549.
- [31] Zhang H, Chen W, He H, Jin Y. Disentangled makeup transfer with generative adversarial network. arXiv Preprint. Source: <https://arxiv.org/abs/1907.01144>.
- [32] Kharchevnikova AS, Savchenko AV. Neural networks in video-based age and gender recognition on mobile platforms. Opt Mem Neural Networks 2018; 27(4): 246-259.

Сведения об авторах

Соколова Анастасия Дмитриевна, 1996 года рождения, в 2020 году окончила Национальный исследовательский университет Высшая школа экономики по магистерской программе «Интеллектуальный анализ данных», с 2020 года обучается в аспирантуре Национального исследовательского университета Высшая школа экономики по направлению «Информатика и вычислительная техника», работает стажером-исследователем в лаборатории ЛАТАС НИУ ВШЭ. Область научных интересов: обработка изображений, распознавание лиц. E-mail: adsokolova@hse.ru.

Савченко Андрей Владимирович, 1985 года рождения, в 2008 году окончил Нижегородский государственный технический университет им. Р.Е. Алексеева по специальности «Прикладная математика и информатика». В 2010 году защитил диссертацию на соискание ученой степени кандидата технических наук по специальности 05.13.18 «Математическое моделирование, численные методы и комплексы программ». В 2015 г. присвоено ученое звание доцента по специальности 05.13.18. В 2016 году присуждена учёная степень доктора технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации». В настоящее время работает профессором кафедры информационных систем и технологий и ведущим научным сотрудником лаборатории алгоритмов и технологий анализа сетевых структур в Национальном исследовательском университете Высшая школа экономики – Нижний Новгород. Автор более 100 научных работ. Область научных интересов: обработка мультимедийной информации, распознавание образов. E-mail: avsavchenko@hse.ru.

Николенко Сергей Игоревич, 1984 года рождения, в 2005 году окончил Санкт-Петербургский государственный университет по специальности «Математика». В 2009 году защитил диссертацию на соискание степени кандидата физико-математических наук. В настоящее время работает научным сотрудником лаборатории математической логики Санкт-Петербургского отделения Математического института им. В. А. Стеклова и доцентом факультета математики и компьютерных наук Санкт-Петербургского государственного университета. Область научных интересов: машинное обучение, сетевые алгоритмы. E-mail: snikolenko@gmail.com.

ГРНТИ: 28.23.15

Поступила в редакцию 15 октября 2021 г. Окончательный вариант – 18 апреля 2022 г.

Open-set face identification with automatic detection of out-of-distribution images

A.D. Sokolova¹, A.V. Savchenko¹, S.I. Nikolenko²

¹ National Research University Higher School of Economics,
603093, Nizhny Novgorod, Russia, Rodionova 136,

² Saint Petersburg University, 199034, Saint-Petersburg, Russia, Universitetskaya nab. 7-9

Abstract

One of main issues in face identification is the lack of training data of specific type (bad quality image, varying scale or illumination, children/old people faces, etc.). As a result, the recognition accuracy may be low for input images which are not similar to the majority of images in the dataset used to train the feature extractor. In this paper, we propose that this issue is dealt with by the automatic detection of such out-of-distribution data based on the addition of a preliminary stage of their automatic rejection using a special convolutional network trained using a set of rare data collected using various transformations. To increase the computational efficiency, the decision about the presence of a rare image is made on the basis of the same face descriptor that is used in the classifier. Experimental research confirmed the accuracy improvement of the proposed approach for several datasets of faces and modern neural network descriptors.

Keywords: face recognition, anomaly detection, image processing, detection of out-of-distribution images.

Citation: Sokolova AD, Savchenko AV, Nikolenko SI. Open-set face identification with automatic detection of out-of-distribution images. *Computer Optics* 2022; 46(5): 801-807. DOI: 10.18287/2412-6179-CO-1061.

Acknowledgements: The research was supported by the Russian Science Foundation under RSF grant 20-71-10010. The research work by S. Nikolenko was supported by St. Petersburg State University under project # 73555239 "Artificial Intelligence and Data Science: Theory, Technology, Industrial and Interdisciplinary Research and Applications".

Authors' information

Anastasiia Dmitrievna Sokolova (b. 1996) graduated from National Research University Higher School of Economics in 2020, master degree in "Data Mining". Since 2020 she has been studying in the postgraduate course of the National Research University Higher School of Economics in the field of "Informatics and Computer Science". Currently she works as intern in the laboratory LATNA. Research interests are image processing, face recognition. E-mail: adsokolova@hse.ru.

Andrey Vladimirovich Savchenko, (b. 1985), graduated from N. Novgorod State Technical University in 2002, majoring in Applied Mathematics and Informatics. He defended his PhD in Mathematical Modeling, Numeric Methods and Software Complexes in 2010. He received the Doctor of Science degree in System Analysis, Control and Information Processing in 2016. Currently he works as the professor of Information Systems and Technologies department and leading researcher of the laboratory of Algorithms and Technologies in Network Analysis in National Research University Higher School of Economics, Nizhny Novgorod. He is the co-author of more than 100 scientific papers. Research interests include multimedia processing and pattern recognition. E-mail: avsavchenko@hse.ru.

Sergey Igorevich Nikolenko, (b. 1984), graduated from St. Petersburg State University in 2005 with a degree in Mathematics. In 2009 he defended his PhD in Physical and Mathematical Sciences. Currently he works as the researcher at the Laboratory of Mathematical Logic of the St.-Petersburg Department of the Steklov Mathematical Institute, and the Associate Professor in Saint Petersburg University. E-mail: snikolenko@gmail.com.

Received October 15, 2021. The final version – April 18, 2022.
