ИЗВЕСТИЯ РАН. ТЕОРИЯ И СИСТЕМЫ УПРАВЛЕНИЯ, 2020, № 4, с. 103–109

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ

УДК 004.852, 004.932

ОБУЧЕНИЕ МУЛЬТИМОДАЛЬНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ОПРЕДЕЛЕНИЯ ПОДЛИННОСТИ ИЗОБРАЖЕНИЙ[1]

© 2020 г. О. В. Гринчук[a,*], В. И. Цурков[b,]**

[a] *МФТИ, Долгопрудный, МО, Россия*

[b] *ФИЦ ИУ РАН, Москва, Россия*

**e-mail: oleg.grinchuk@phystech.edu*

***e-mail: tsurkov@ccas.ru*

Поступила в редакцию 06.02.2020 г.
После доработки 28.02.2020 г.
Принята к публикации 30.03.2020 г.

Определение попыток подмены изображений играет важную роль в защите систем биометрии (авторизация в мобильных устройствах, системы контроля и управления доступом в помещения, терминалы с автоматическим доступом по лицу и т.д.). Представлен новый метод для детектирования фальсифицированных изображений, основанный на обработке мультимодальных данных с камеры. Разработана новая архитектура нейронной сети, агрегирующая признаки с разных модальностей на всех уровнях модели. Рассмотрено разделение тренировочной выборки по разным типам атак и инициализация модели признаками, обученными на других задачах, которые связаны с изображениями лиц. Проведены численные эксперименты на реальных данных, показывающие успешную работоспособность системы. Предложенная модель заняла первое место на конкурсе по распознаванию поддельных изображений лиц CASIA-SURF.

DOI: 10.31857/S0002338820040071

Введение. Вместе со стремительным развитием технологий биометрической аутентификации появилась острая необходимость в защите от попыток обхода системы распознавания лиц. Прежде чем отправлять биометрические образцы на процедуру верификации личности, защитная система должна уметь определять, кто стоит перед камерой – живой человек или сфальсифицированный объект: распечатанная фотография, проигрывание видеозаписи с экрана устройства, силиконовые трехмерные маски и другие методы взлома.

Современные системы распознавания лиц уже превосходят способности человека в этой области [1]. Значительная часть такого успеха обусловлена наличием больших размеченных датасетов [2, 3], обычно собираемых из Интернета. В отличие от задачи распознавания лиц, датасеты для задачи определения подлинности (liveness) требуют тщательного ручного сбора данных, так как изображений различных фальсификаций нет в свободном доступе. Такие данные собираются в лабораториях с приглашенными участниками и, следовательно, сильно ограничены в количестве и разнообразии, что нивелирует все преимущества нейросетей при работе с обычными цветными изображениями. С другой стороны, для определения liveness можно использовать не только обычные камеры, но и специальные сенсоры, предоставляющие дополнительные модальности для анализа. Эти модальности добавляют дополнительную информацию и могут улучшить детектирование liveness. Например, инфракрасная (ИК) камера нечувствительна к экранам электронных устройств и автоматически защищает от возможных подлогов такого типа. Камера глубины позволяет получить трехмерное изображение объекта, делая нахождение любых плоских (отличающихся от формы лица) подделок проще.

Большинство датасетов антиспуфинга лиц содержат только изображения в формате RGB (red, green, blue – аддитивная цветовая модель, описывающая способ кодирования цвета) [4, 5]. До недавнего времени датасеты, содержащие другие модальности, были очень ограничены в количестве примеров [6, 7], что увеличивало риск переобучения модели на тренировочную выборку.

[1] Работа выполнена при частичной финансовой поддержке РФФИ (грант № 19-01-00625).

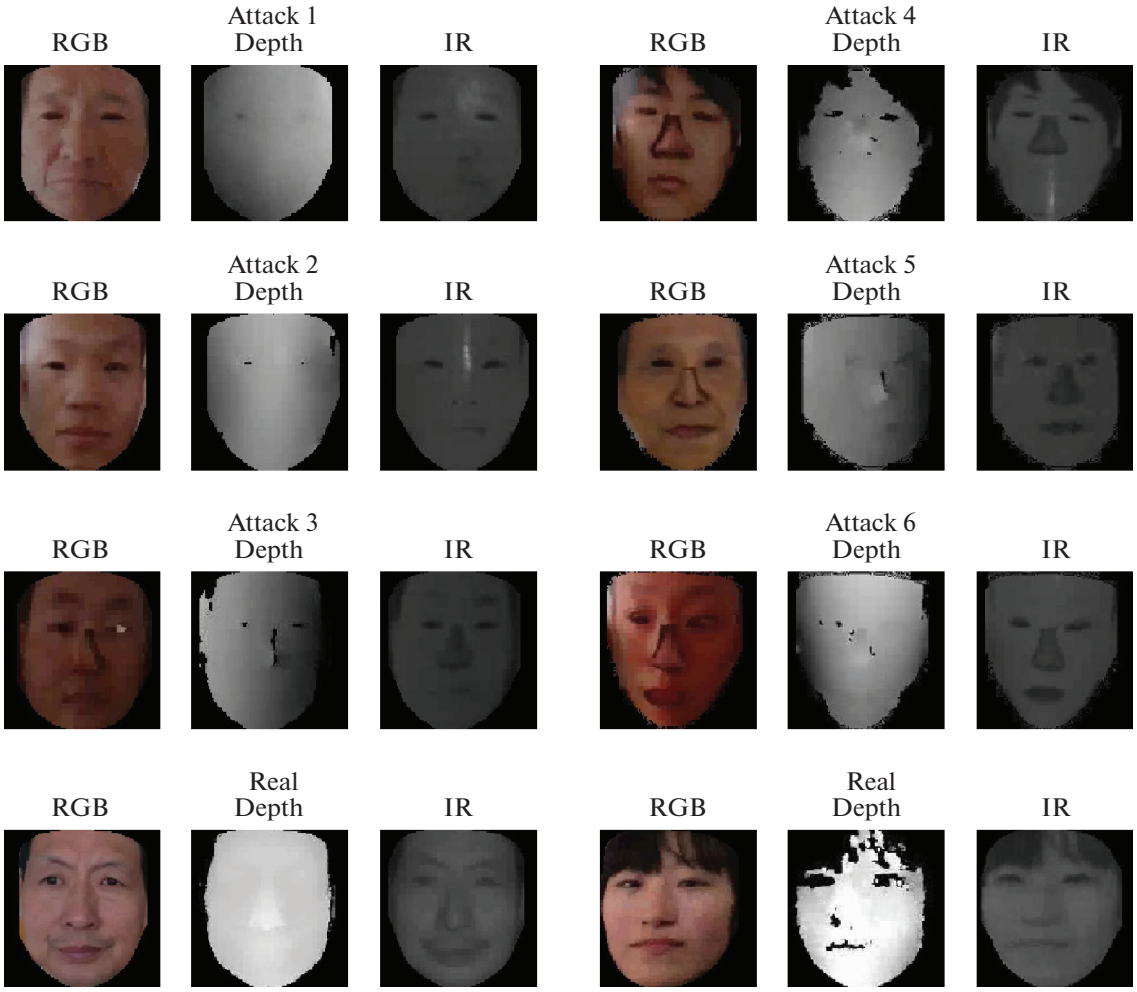

Рис. 1. Примеры реальных и поддельных изображений из датасета CASIA-SURF

Недавно выпущенный liveness датасет CASIA-SURF [8] на порядок лучше предыдущих с точки зрения как количества данных, так и количества доступных модальностей (RGB, ИК, глубина), что позволяет эффективно применить инструментарий нейронных сетей для решения задачи liveness.

В данной статье предлагается новый метод решения задачи liveness, основанный на модификации архитектуры сети из [8]. Мы обрабатываем каждую модальность отдельно, при этом объединяя признаки на разных уровнях глубины сети, что увеличивает обмен полезными признаками между каналами глубины, RGB и ИК.

Несмотря на большое количество изображений, CASIA-SURF все равно на порядки меньше стандартных датасетов для задачи распознавания лиц. Для обучения глубоких сверточных нейронных сетей этого недостаточно. При этом обе задачи похожи, так как работают с изображениями лиц. Поэтому предлагается инициализировать веса нейронной сети для задачи liveness весами модели из задачи распознавания и потом дообучать модель на меньшем датасете liveness. В таком случае в начале обучения на целевой задаче параметры модели уже находятся в окрестности оптимальных, включая в себя сильные признаки изображений лиц, что увеличивает вероятность получить итоговую хорошую модель. В данной работе используются четыре начальные сети, обученные на разных датасетах распознавания и определения атрибутов лиц, которые стабилизируют финальную модель liveness и увеличивают ее точность.

Для увеличения устойчивости против неизвестных методов взлома обучается несколько сетей на разных подмножествах обучающей выборки, содержащих разные типы атак. Итоговая модель получается путем усреднения предсказаний всех нейросетей. В результате предложенный метод

достигает 99.8739 TPR в FPR = 10^{-4} на тестовой выборке CASIA-SURF, заняв первое место в соревновании “Chalearn LAP multi-modal face anti-spoofing attack detection challenge” [9].

1. Обзор литературы. Есть два общих подхода к решению задачи противодействия мошенничеству в биометрических системах безопасности. Первое семейство методов [10–12] предполагает кооперативное взаимодействие с пользователем, запрашивая определенные движения, такие, как моргание, кивки, улыбка и другие изменения положения головы или мимики. Эти методы обеспечивают хорошую защиту от атак с помощью распечатанных изображений и фиксированных трехмерных масок, но беззащитны перед демонстрацией видеозаписи с нужными действиями. Более того, интерактивные методы применимы не во всех сценариях и часто неудобны и раздражительны для пользователя системы.

Второе семейство методов [13–16] нацелено на определение liveness по одному кадру лица. Такие алгоритмы обычно работают быстро и незаметно для пользователя, но при этом сложнее получить требуемую точность, особенно если используются только RGB-изображения ввиду отсутствия большого разнообразного датасета и ограничений модальности видимого спектра.

Есть несколько датасетов, которые можно рассматривать для разработки некооперативного определения liveness. Датасеты Replay-Attack [17], CASIA-FASD [18] и SiW [16] состоят из RGB-фотографий. MSU-MFSD [19], Replay-Mobile [5] и OULU-NPU [4] содержат видеозаписи атак с мобильных устройств. Но из-за недостаточного количества данных модели, обученные на этих данных плохо обобщаются к другим условиям съемки и видам камер. Более того, с развитием методов взлома (реалистичные трехмерные силиконовые маски) и увеличением разрешения экранов мобильных устройств одного RGB-канала становится недостаточно для создания решения с высокой точностью. Добавление дополнительных модальностей, обладающих полезными признаками, позволяет решить эту проблему.

2. Датасет CASIA-SURF. Датасет CASIA-SURF [8] включает в себя 21000 видеозаписей 1000 субъектов, для каждого субъекта записано одно реальное видео и шесть поддельных видео, содержащих разные виды атак с лицом этого человека. Видео записаны с помощью камеры Intel RealSense SR300 и имеют три синхронизированных канала: RGB, ИК, глубина. Выборка разделена на обучающую, валидационную и тестовую подвыборки, содержащие 300, 100 и 600 уникальных субъектов соответственно. Из каждого видео выбран каждый десятый кадр, переводя датасет в набор изображений. Кроме того, выборки были также разделены по типам атак, в тестовой выборке присутствуют фальсификации, которых не было в обучающей выборке. После релиза датасета авторы статьи запустили соревнование на лучшее решение для тестовой части, сделав доступными 40000 изображений для обучения и валидации.

Примеры настоящих и поддельных изображений из CASIA-SURF показаны на рис. 1. Атаки отличаются формой (плоская, согнутая) и вырезанными частями лица (табл. 1) для создания объемности подделке. Атаки, представленные в тестовой части, полностью отличаются от содержащихся в обучающей выборке. В таком разбиении данных для демостратции высокой точности модель должна обладать обобщающей способностью и избегать переобучения на конкретные виды атак, что являлось большой проблемой в ранее опубликованных датасетах.

2.1. Базовый метод. Вместе с релизом CASIA-SURF [8] авторы также предложили базовый метод решения. Нейронная сеть обрабатывает каждую из модальностей отдельно, используя архитектурные блоки из resnet-18 [20] в качестве основы. Далее совершается перебалансировка признаков каждой ветви, выбираются наиболее информативные признаки и подавляются остальные. Выходы с каждой из трех ветвей объединяются в один и обрабатываются еще двумя resnet-блоками. Завершают архитектуру глобальный слой усреднения и два полносвязных слоя.

Таблица 1. Виды фальсификационных атак из CASIA-SURF

Поверхность	Глаза	Нос	Рот	Выборка
Плоская	✓			Тест
Согнутая	✓			»
Плоская	✓	✓		»
Согнутая	✓	✓		»
Плоская	✓	✓	✓	Обучение
Согнутая	✓	✓	✓	»

Авторы провели тщательные эксперименты и показали преимущества предложенной модели, в данной статье мы применяем ее в качестве стартовой точки.

2.2. Метрики оценивания точности. Существуют различные метрики оценивания точности алгоритма определения liveness. Одной из популярных метрик является average classification error rate (ACER), используемая в работах [4, 5, 21, 22]. Авторы CASIA-SURF предлагают основную метрику распознавания лиц – true positive rate (TPR) в зафиксированном false positive rate (FPR). Такой подход позволяет оценить, сколько реальных пользователей пройдут тест на попытку взлома, при этом пропуская только определенный процент атак. Мы рассматриваем TPR в 10^{-4} FPR, которую можно получить из ROC-кривой (receiver operating characteristic) на целевой выборке.

3. Предлагаемый метод. В данном разделе описаны детали рассматриваемого метода.

3.1. Разбиение обучающей выборки. В CASIA-SURF атаки, представленные в обучающей выборке, отличаются от тестовых атак. Для увеличения устойчивости модели к новым атакам мы выделили из обучающей выборки три части. Каждая часть содержит все изображения двух разных атак, данные по третьей атаке используются как валидационная выборка. После чего обучаются три нейронные сети на каждой из частей. Во время тестирования, все модели рассматриваются как одна, выходы с классификационного слоя усредняются по трем значениям выходов каждой из обученной сети.

3.2. Перенос признаков. Множество задач компьютерного зрения [23–25] с небольшим доступным объемом данных для обучения в качестве инициализации применяют обученные модели других задач, в которых выборка достаточно большая [26]. Дообучение параметров сети, которая была инициализирована предобученными параметрами разных задач, приводит к различным результатам на тестовом сете. В наших экспериментах мы тестируем четыре разные модели, предобученные на разных датасетах распознавания лиц и классификации пола. Кроме того, в этих задачах мы использовали разные архитектуры базовой модели и функции потерь для увеличения вариативности итоговых параметров. После дообучения на задаче liveness четыре итоговые модели применяются как одна путем усреднения предсказаний.

3.3. Архитектура модели. Предлагаемая архитектура основана на resnet-34 и resnet-50 с SE-модулями (squeeze and excitation) [20], как показано на рис. 2. Следуя протоколу, описанному в [8], каждая модальность обрабатывается первыми тремя блоками архитектуры resnet, дальше три ветви объединяются с помощью SE-модуля и обрабатываются оставшимся res-блоком. В отличие от базового метода мы обогатили модель дополнительными блоками агрегации на каждом слое ветвей (мультиуровневая агрегация признаков – МУАП). Агрегационный блок берет признаки с соответствующих res-блоков подсетей модальностей и из предыдущего агрегационного блока и обрабатывает их. Такая архитектура позволяет нейронной сети находить корреляцию между признаками не только высокого, но и низкого уровня.

4. Эксперименты. В данном разделе описываются технические детали решения, а также показывается влияние каждого из предложенных улучшений на целевую метрику качества.

4.1. Технические детали. Код был написан с помощью библиотеки Pytorch [27], нейронные сети обучались на четырех видеокартах NVIDIA 1080Ti. Обучение одной модели занимает 3 ч, обученная модель извлекает предсказания по 1000 изображениям за 8 с.

Все нейронные сети были обучены с помощью оптимизатора ADAM [28], параметр скорости обучения изменялся по косинусу, в качестве функции потерь использовалась стандартная двухклассовая кроссэнтропия. Модель обучалась 30 эпох с начальным параметром скорости обучения 0.1 и размером минибатча 128. Эти же параметры применялись для обучения моделей распознавания лиц.

4.2. Предобработка. Изображения в датасете CASIA-SURF уже вырезаны по контуру лица, поэтому никакого дополнительного выравнивания лиц не потребовалось. Изображения изменялись до 125 × 125 пикселей, после чего вырезался центральный регион размером 112 × 112. В процессе обучения картинки случайно отзеркаливались по горизонтали с вероятностью 0.5. Также были проверены другие стратегии предобработки, но они не принесли улучшений по сравнению с описанной.

4.3. Базовый метод. Датасет разбит на обучающую, валидационную и тестовую выборки, но так как в момент соревнования Chalearn LAP тестовая часть была недоступна, далее все результаты приводятся для валидационной части. В первую очередь мы воспроизвели базовый метод из [8], основанный на resnet-18 и обучили пять сетей на пяти частях по стратегии

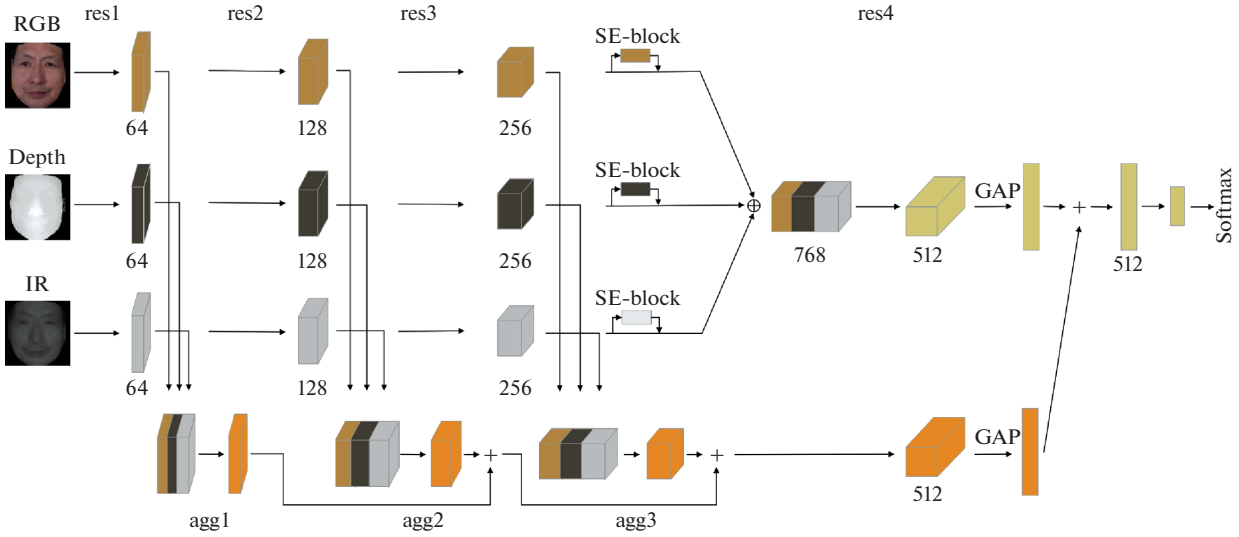

Рис. 2. Предлагаемая архитектура. GAP – общий слой усреднения; ⊕ – оператор объединения; + – оператор почленного суммирования

кроссвалидации. Все части были разделены по субъектам, все изображения одного субъекта принадлежали только одной части. Итоговая модель – результат усреднения предсказаний пяти полученных моделей. В табл. 2 приведены результаты из статьи базового метода и результаты нашего воспроизведения.

Далее, мы расширили основу сети до resnet-34, что сильно увеличило точность на целевой выборке. Ввиду ограничений вычислительных ресурсов мы обучали только модели resnet-34 и resnet-50, не тестируя более глубокие сети.

4.4. Разбиение обучающей выборки. В данном эксперименте сравниваются результаты моделей, обученных стандартным методом кроссвалидации по пяти частям и предложенным методом разбиения обучающей выборки на три части по типам атак. Изображения реальных людей в таком разбиении случайно разделены по этим частям. Несмотря на то, что новая модель получена усреднением трех сетей, а не пяти, которые к тому же обучены на меньшем количестве данных, чем в стандартном разбиении, ее результаты лучше на 4.3% (табл. 2). Это может быть объяснено тем, что разбиение по типам атак в обучающей выборке позволяет лучше адаптироваться к неизвестным примерам фальсификации из целевого сета.

Таблица 2. Результаты на валидационной выборке CASIA-SURF

Метод	Инициализация	Обучающая выборка	TPR at FPR=10^{-4}
Zhang, Wang et al. [8]	Нет	Одна часть	56.80
resnet-18	»	Пять частей по субъектам	60.54
resnet-34	»	»	74.55
resnet-34	»	Три части по атакам	78.89
resnet-34	ImageNet [26]	»	92.12
resnet-34	CASIA-Web face [29]	»	99.80
A. resnet-34 + МУАП	CASIA-Web face [29]	»	99.87
B. resnet-50 + МУАП	MSCeleb-1M [2]	»	99.63
C. resnet-50 + МУАП	Asian dataset [30]	»	99.33
D. resnet-34 + МУАП	AFAD-lite [31]	»	98.70
Усреднение A,B,C,D	–	»	100.00

Таблица 3. Влияние дополнительных модальностей на целевую метрику

Модальность	TPR в FPR в точках		
	10^{-2}	10^{-3}	10^{-4}
RGB	71.74	22.34	7.85
ИК	91.82	72.25	57.41
глубина	100.00	99.77	98.40
RGB+ИК+глубина	100.00	100.00	99.87

4.5. Инициализация весов. В текущем разделе мы исследуем зависимость целевой точности от задания начальных параметров. Параметры каждой из трех ветвей архитектуры инициализируются весами сети, обученной на ImageNet [26], после чего дообучаются на CASIA-SURF. В сравнении со случайной инициализацией, применение предобученной сети увеличивает результат с 78.89 до 92.12%.

Если же вместо общего датасета классификации изображений ImageNet использовать датасет для задачи распознавания лиц CASIA-Web [29], точность достигает почти идеального значения 99.80%

4.6. МУАП. При дополнении стандартной архитектуры блоком предложенной в данной работе МУАП новая модель после обучения показывает уменьшение ошибки в 1.5 раза по сравнению с базовой моделью (табл. 2).

4.7. Ансамбль моделей. Для улучшения устойчивости решения используются четыре модели, предобученные на четырех различных датасетах: A. CASIA-WebFace [29], B. MSCeleb-1M [2], C. AsianDataset [30] и D. AFAD-lite [31]. Разные исходные задачи, датасеты и функции потерь приводят с разным обученным весам сверточных фильтров, в итоге финальная модель как усреднение сетей A, B, C и D позволяет достичь 100.00% TPR в FPR = 10^{-4} (табл. 2).

4.8. Мультимодальность. Чтобы показать преимущество мультимодальных данных в задаче определения liveness, мы исследовали сети, обученные только на одной модальности. Для честного сравнения использовалась та же архитектура, что и для мультимодальных изображений, только вместо подавания на вход (RGB, ИК, глубина), модели обучались на входах (RGB, RGB, RGB), (ИК, ИК, ИК) и (глубина, глубина, глубина).

Как видно в табл. 3, использование только одного канала RGB приводит к низкой точности. Соответствующая модель переобучилась на тренировочной выборке и достигла только 7.85% TPR в FPR = 10^{-4}. Модель на инфракрасных данных оказалась лучше, показав 57.41% TPR в FPR = 10^{-4}. ИК-данные содержат меньше мелких деталей, поэтому сети, основанные на них, сложнее переобучаются и в общем более устойчивы на неизвестных данных, что и показал текущий результат. Наиболее высокий результат 98.40% TPR в FPR = 10^{-4} был получен на модальности глубина, подтвердив важность информации о форме для задачи проверки подлинности лица.

Но сеть, обученная на объединении модальностей, показала еще лучшую точность, понижая ложноотрицательную ошибку с 1.6 до 0.13% и доказывая важность мультимодального подхода.

Заключение. В данной работе был представлен новый метод для решения задачи детектирования фальсифицированных изображений лиц, который показал лучший результат на конкурсе “Chalearn LAP face anti-spoofing”. Были в деталях предложены три направления работы: данные, архитектура нейронной сети и инициализация весов. Комплексный подход выявил существенные улучшения точности по сравнению с базовым методом. Тщательный выбор обучающей подвыборки по типам атак позволяет модели лучше противостоять незнакомым попыткам взлома. Предложена новая архитектура сети с модулем мультиуровневой агрегации признаков, что улучшило обмен полезными признаками между подсетями разных модальностей как на поверхностных, так и на глубоких слоях модели. Использован метод переноса признаков с обученных моделей распознавания лиц, что улучшило стабильность модели и увеличило точность на целевой выборке.

СПИСОК ЛИТЕРАТУРЫ

1. *Phillips J., Yates A., Hu Y. et al.* Face Recognition Accuracy of Forensic Examiners, Superrecognizers, and Face Recognition Algorithms // Proc. National Academy of Sciences. 2018. V. 15. P. 6171–6176.
2. *Guo Y., Zhang L., Hu Y., He. X, Gao J.* MS-Celeb-1M: A Dataset and Benchmark for Large Scale Face Recognition // European Conf. on Computer Vision. Amsterdam, 2016.

3. *Parkhi O., Vedaldi A., Zisserman A.* Deep Face Recognition // British Machine Vision Conf. Swansea, UK, 2015.
4. *Boulkenafet Z., Komulainen J., Li L., Feng X., Hadid A.* Oulu-npu: A Mobile Face Presentation Attack Database with Real-world Variations // Conf. on Automatic Face and Gesture Recognition. Washington, DC, 2017.
5. *Costa-Pazo A., Bhattacharjee S., Vazquez-Fernandez E., Marcel S.* The Replay-Mobile Face Presentation-attack Database // Proc. Int. Conf. on Biometrics Special Interests Group (BioSIG). Darmstadt, 2016.
6. *Chingovska I., Erdogmus N., Anjos A., Marcel S.* Face Recognition Systems under Spoofing Attacks // Face Recognition Across the Imaging Spectrum. Springer, Cham, 2016.
7. *Erdogmus N., Marcel S.* Spoofing in 2D Face Recognition with 3D Masks and Anti-spoofing with Kinect // IEEE Sixth Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS). Arlington, VA, 2014.
8. *Zhang S., Wang X., Liu A. et al.* A Dataset and Benchmark for Large-scale Multi-modal Face Anti-spoofing // CVPR. Long Beach, CA, 2019.
9. *Liu A., Wan J., Escalera S. et al.* Multi-modal Face Anti-spoofing Attack Detection Challenge at CVPR2019 // CVPR Workshop. Long Beach, CA, 2019.
10. *Pan G., Sun L., Wu Z., Lao S.* Eyeblink-based Anti-spoofing in Face Recognition from a Generic Webcamera // Int. Conf. on Computer Vision. Venice, 2007.
11. *Wang L., Ding X., Fang C.* Face Live Detection Method Based on Physiological Motion Analysis // Tsinghua Science and Technology. 2009. V. 14. P. 685–690.
12. *Bharadwaj S., Dhamecha T., Vatsa M., Singh R.* Computationally Efficient Face Spoofing Detection with Motion Magnification // CVPR Workshop. Portland, 2013.
13. *Feng L., Po L., Li Y. et al.* Integration of Image Quality and Motion Cues for Face Antispoofing: A Neural Network Approach // J. Visual Communication and Image Representation. 2016. V. 38. P. 451–460.
14. *Patel K., Han H., Jain A.* Secure Face Unlock: Spoof Detection on Smartphones // Transactions on Information Forensics and Security. 2016. V. 11. P. 2268–2283.
15. *Li L., Feng X., Boulkenafet Z., Xia Z., Li M., Hadid A.* An Original Face Antispoofing Approach Using Partial Convolutional Neural Network // Sixth Int. Conf. on Image Processing Theory, Tools and Applications (IPTA). Oulu, 2016.
16. *Liu Y., Jourabloo A., Liu X.* Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision // CVPR. Salt Lake City, 2018.
17. *Chingovska I., Anjos A., Marcel S.* On the Effectiveness of Local Binary Patterns in Face Antispoofing // Proc. Int. Conf. on Biometrics Special Interests Group (BioSIG). Darmstadt, 2012.
18. *Zhang Z., Yan J., Liu S. et al.* A Face Antispoofing Database with Diverse Attacks // International Conf. on Biometrics. New Delhi, 2012.
19. *Wen D., Han H., Jain A.* Face Spoof Detection with Image Distortion Analysis // Transactions on Information Forensics and Security. 2015. V. 10. P. 746–751.
20. *He K., Zhang X., Ren S., Sun J.* Deep Residual Learning for Image Recognition // CVPR. Las Vegas, 2016.
21. *Liu Y., Jourabloo A., Liu X.* Learning Deep Models for Face Anti-Spoofing: Binary or Auxiliary Supervision // CVPR. Salt Lake City, 2018.
22. *Jourabloo A., Liu Y., Liu X.* Face Despoofing: Anti-spoofing via Noise Modeling // European Conf. on Computer Vision. Munich, 2018.
23. *Визильтер Ю.В., Желтов С.Ю.* Использование проективных морфологий в задачах обнаружения и идентификации объектов на изображениях // Изв. РАН. ТиСУ. 2009. № 2. Р. 125–138.
24. *Кузнецов В.Д., Матвеев И.А., Мурынин А.Б.* Идентификация объектов по стереоизображениям. II. Оптимизация информационного пространства // Изв. РАН. ТиСУ. 1998. № 4. С. 50–53.
25. *Соломатин И.А., Матвеев И.А., Новик В.П.* Определение видимой области радужки классификатором текстур с опорным множеством // АиТ. 2018. № 3. С. 127–143.
26. *Deng J., Dong W., Socher R., Li L.-J., Li K., Fei-Fei L.* ImageNet: A Large-Scale Hierarchical Image Database // CVPR. Miami, 2009.
27. *Paszke A., Gross S., Chintala S. et al.* Automatic Differentiation in PyTorch // NIPS workshop. Long Beach, CA, 2017.
28. *Kingma D., Ba J.* Adam: A Method for Stochastic Optimization // Int. Conf. on Learning Presentations. San Diego, 2015.
29. *Yi D., Lei Z., Liao S., Li S.* Learning Face Representation from Scratch // arXiv. 2014. http://arxiv.org/abs/1411.7923.
30. *Zhao J., Cheng Y., Xu Y. et al.* Towards Pose Invariant Face Recognition in the Wild // CVPR. Salt Lake City, 2018.
31. *Niu Z., Zhou M., Wang L., Gao X., Hua G.* Ordinal Regression With Multiple Output CNN for Age Estimation // CVPR. Las Vegas, 2016.