Обнаружение спуфинг-атак на систему лицевой биометрии

И.А. Калиновский[1,2], Г.М. Лаврентьева[2,3]
kalinovskiy@speechpro.com|lavrentyeva@speechpro.com
[1]ООО «ЦРТ», Санкт-Петербург, Российская Федерация;
[2]ООО «ЦРТ-инновации», Санкт-Петербург, Российская Федерация;
[3]Университет ИТМО, Санкт-Петербург, Российская Федерация

В статье предлагается способ обнаружения спуфинг-атак на систему лицевой биометрии, основанный на глубоких нейронных сетях. Разработанная мультимодальная архитектура сверточной нейронной сети с механизмом внимания, позволяет достичь лучших результатов на бенчмарке CASIA-FASD, превосходя другие нейросетевые модели по качеству детектирования подложных образцов. Приводятся результаты ряда экспериментов по исследованию различных алгоритмов предобработки изображения для улучшения качества детектирования спуфинга. Показано, что использование метода нормализации Tan-Triggs позволяет значительно улучшить качество работы нейросетевого классификатора для рассматриваемой задачи.

Ключевые слова: *спуфинг-атака, биометрические системы, сверточные нейронные сети.*

Face anti-spoofing for biometric systems

I.A. Kalinovskiy[1,2], G.M. Lavrentyeva[2,3]
kalinovskiy@speechpro.com|lavrentyeva@speechpro.com
[1]"STC" Ltd., Saint Petersburg, Russian Federation;
[2]"STC-Innovations" Ltd., Saint Petersburg, Russian Federation;
[3]ITMO University, Saint Petersburg, Russian Federation

This paper proposes spoofing detection approach for face recognition systems, based on deep neural networks. The suggested architecture of multimodal convolutional neural network with attention mechanism allows to achieve state-of-the-art results for the CASIA-FASD benchmark database, exceeding other neural network models in terms of spoofing-detection quality. This research presents experimental results of the investigation of different image preprocessing techniques implemented to increase spoofing detection quality. Results confirm that Tan-Triggs normalization method leads to significant quality improvement of neural network classifier for the considered task.

Keywords: *spoofing attack, biometric systems, convolutional neural networks.*

1. Введение

Биометрические системы распознавания личности в последнее время получили широкое распространение. В основе подобных систем лежат технологии преобразования различных типов индивидуальных биометрических характеристик (отпечатки пальцев, радужная оболочка глаза, голос, лицо и др.) в цифровой код, который используется для решения задач верификации и идентификации профиля человека среди миллионов записей в базе данных. Наиболее простой с точки зрения считывания биометрических дынных является технология лицевой биометрии [11]. Современные системы распознавания личности по лицу способны с высокой точностью работать даже в неконтролируемых условиях. Однако это приводит к повышенному риску взлома, так как для прохождения верификации, системе достаточно предоставить фотографию, сделанную на обычную камеру или взятую из открытых источников. В связи с этим возникает ряд задач по предотвращению попыток подмены биометрических данных, обычно называемых спуфинг-атаками.

Спуфинг-атака на лицевую биометрическую систему может быть осуществлена разными способами. Наиболее эффективным из них является прямая загрузка фотографии в систему, однако для этого злоумышленникам требуется получить доступ к программному обеспечению, что значительное усложняет атаку. В случае доступности ответа системы в виде степени схожести или вероятности верификации, возможно осуществление атаки непосредственно на алгоритмы построения биометрического шаблона, которые, как правило, основаны на глубоких сверточных нейронных сетях (СНС) [11]. Как известно СНС уязвимы к так называемым «состязательным атакам» (adversarial attack) [21], что позволяет злоумышленникам использовать специальным образом нанесенный макияж для существенного увеличения вероятности ложного срабатывания системы. Наконец, наиболее простым способом осуществления атаки является атака на уровне сенсора с использованием фотографии зарегистрированного в системе пользователя.

В данной работе предлагается способ защиты от последнего вида спуфинг-атак на системы контроля и управления доступом (СКУД). Ввиду большой популярности социальных сетей весьма просто заполучить фотографию какого-либо человека, работающего на предприятии. При этом она может демонстрироваться сенсору как в распечатанном виде, так и на смартфоне или планшете (рис. 1). Далее будет рассмотрена единая модель, позволяющая детектировать одновременно все варианты спуфинга на уровне сенсора используя только одно изображение при малом количестве ложных отказов.

Рис 1. Примеры спуфинг-атак из базы CASIA-FASD [22]

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

2. Обзор методов обнаружения спуфинга

Проблема детектирования спуфинг-атак или обнаружения витальности (liveness detection) применительно к системам лицевой биометрии, приобрела популярность у исследователей в середине 2000-х готов [15]. К настоящему времени предложено множество методов ее решения, которые условно разделяются на две группы: активные и пассивные. Активные методы запрашивают от пользователя совершения определенного действия, например: улыбнуться, моргнуть, наклонить или повернуть голову и др. [2, 18]. Пассивные методы обычно используют для анализа только одного изображения, по которому непосредственно строится биометрический шаблон [17]. В связи с этим они удобнее в использовании, а также позволяют исключить ситуацию демонстрации фотографии в промежуток между процессами определения витальности лица и верификации.

Большое количество работ, посвященных алгоритмам пассивного обнаружения спуфинг-атак, основаны на анализе текстуры области лица. Для описания текстуры обычно используются различные разновидности LBP [5], SURF [3], HOG-LPQ [12] и других дескрипторов, а в качестве классификатора – SVM. При этом исследуется влияние различных цветовых пространств на точность обнаружения, а также применяется анализ в частотной области [8]. Несмотря на хорошее качество подобных алгоритмов, достигаемое на стандартных бенчмарках [22], оно существенно ухудшается при изменении условий регистрации изображений.

В ряде работ изучается возможность использования оптических эффектов, создаваемых камерой (муар [13], расфокусировка [7], дисторсия [6]). Недостатком подобных методов является необходимость в точной калибровке алгоритмов для конкретного оборудования. В тоже время использование дополнительного специализированного оборудования, например камер светового поля или 3D-сенсоров, позволяет добиться наиболее высокой степени защиты от спуфинк-атак, но ценой удорожания стоимости СКУД для конечного пользователя.

В последние годы для решения рассматриваемой задачи стали активно применяться методы глубокого обучения. В [20] использовалась СНС типа AlexNet для бинарной классификации реальное лицо/спуфинг-атака, а также исследовалось влияние размера контекста, захватываемого при выравнивании изображения лица, на качество такой классификации. Авторы [9] использовали SVM для классификации высокоуровневых признаков, полученных с последних слоев дообученной сети VGG-Face. В [19] использовалась комбинация СНС и рекуррентной сети типа LSTM для классификации последовательности кадров. В [1] была предложена оригинальная комбинация «patched-based» СНС, классифицирующей группу локальных участков изображения, и «depth-based» СНС, которая генерировала плотную карту глубины для каждого пикселя. Для обучения «depth-based» модели карта глубины реальных изображений лиц генерировалась с помощью нежесткой 3D-модели, а для поддельных задавалась как нулевая. В [10] авторы распространили предложенный подход на обработку последовательности кадров [19].

Несмотря на значительный прогресс в обнаружении витальности лица, эта задача по-прежнему остается не решенной в общем случае. Сверхвысокая плотность пикселей и естественная цветопередача современных дисплеев делают отображаемое изображение лица почти не отличимым от реального. В связи с этим, существует потребность в совершенствовании предложенных алгоритмов обнаружения спуфинга и в разработке новых подходов.

3. Мультимодальная сверточная нейронная сеть с механизмом внимания

В настоящее время глубокие сверточные нейронные сети являются стандартным блоком практически во всех задачах, связанных с обработкой изображений. Современные библиотеки для машинного обучения значительно ускоряют разработку новых нейросетевых архитектур, что приводит к постепенному усложнению их вычислительного графа. Недавние работы в области глубокого обучения продемонстрировали эффективность механизма внимания для улучшения работы СНС в задачах распознавания образов, генерации подписей к изображениям и др. [14]. Блоки внимания позволяют выделять наиболее информативные участки карт признаков на разных уровнях нейронной сети, специфичных для каждого конкретного класса. В основе предлагаемого подхода к обнаружению спуфинг-атак лежит сверточная нейронная сеть с модулями внимания, имеющая несколько ветвей для обработки различных входных данных (таблица 1).

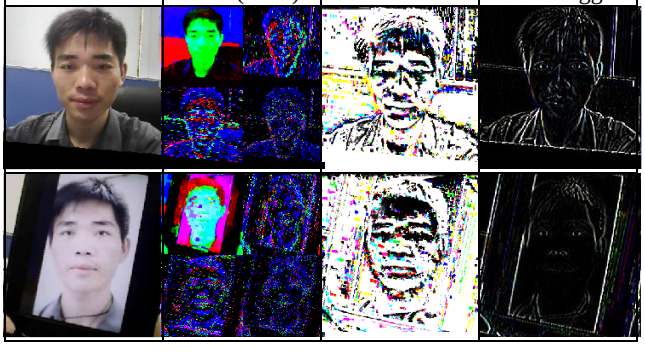
3.1 Мультимодальная архитектура

Для обучения сверточных нейронных сетей обычно используются непосредственно RGB-каналы изображения. В тоже время во множестве предшествующих работ по обнаружению витальности лица была продемонстрирована эффективность использования различных текстурных дескрипторов для кодирования лицевой области. Необходимость обобщать различия в освещенности и позе лиц для двух визуально похожих классов, одновременно с поиском мелкомасштабных текстурных отличий, усложняет задачу оптимизации нейронной сети в контексте рассматриваемой задачи. В связи этим, в настоящей работе предлагается использовать ряд процедур предобработки исходных изображений с целью извлечения наиболее информативных текстурных характеристик.

Для совместной обработки полученных таким образом различных карт-признаков, предлагается архитектура СНС, имеющая несколько идентичных входов (табл. 2). Обработка сетью информации, поступающей на каждый вход, осуществляется независимо вплоть до определенного уровня глубины, после чего полученные карты конкатенируется в единый пул. При этом каждая ветка может включать один или несколько модулей внимания.

В ходе работы были проведены эксперименты с различными алгоритмами «ручного» выделения признаков: цветовые пространства HSV и YCbCr, LBP-дескрипторы, вельветы Хаара и Добеши, алгоритм нормализации Tan-Triggs [16] (табл. 1). Результаты исследования зависимости качества обнаружения спуфинг-атак от алгоритма предобработки входных изображений обсуждаются в разделе 4.

Таблица 1. Иллюстрация вариантов предобработки данных

RGB	Haar (HSV)	LBP	Tan-Triggs
			

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18

Таблица 2. Архитектура CHC для обнаружения спуфинга

Слой	Фильтр / шаг	Размер выхода
Input		192×192×3 (*N*)
N веток для обработки каждого входа		
Conv-1.1 (SELU, BN)	3×3/1	192×192×32
Conv-1.2	3×3/2	96×96×32
Conv-2.1 (SELU, BN)	3×3/1	96×96×64
Conv-2.2	3×3/2	48×48×64
Conv-3.1 (SELU, BN)	3×3/1	48×48×128
Attention-1 (*K*=4)		1×1×2
Conv-3.2	3×3/2	24×24×128
Основная ветка графа		
Concatenation		24×24×(128×*N*)
Conv-4 (SELU, BN)	3×3/1	24×24×256
Attention-2 (*K*=4)		1×1×2
MaxPooling-1	3×3/2	12×12×256
Conv-5 (SELU, BN)	3×3/1	12×12×512
Attention-3 (*K*=4)		1×1×2
MaxPooling-2	3×3/2	6×6×512
Conv-6 (SELU, BN)	3×3/1	6×6×1024
Dropout-1		
FC-1		1×1×2048
Dropout-2		
FC-2		1×1×2
Add (attentions layers)		1×1×2

3.2 Механизм внимания

В данной работе применятся простой механизм внимания, описанный в [14]. Авторы предложили универсальную архитектуру модуля внимания, который может быть подключен к любому уровню сети. При этом для его обучения не требуются дополнительные метки. В состав модуля входят три основных блока:

1) слой внимания, состоящий из сверточного слоя с одним выходом и ядром 1×1, а также softmax-слой, который производит вероятностную карту;
2) выходной полносвязный слой, генерирующий гипотезу о классе на основе взвешенных с выходом softmax-слоя входных карт признаков;
3) ворота (gates), контролирующие уровень выходного сигнала от данного модуля.

В данной работе используется модифицированная архитектура модуля, представленная в таблице 3. К одному участку нейронной сети может быть подключено одновременно *K* модулей внимания, при этом их выходы суммируются.

Таблица 3. Архитектура модуля внимания

Слой	Фильтр / шаг	Размер выхода
Input		*S*×*S*×*D*
Conv-1	3×3/1	*S*×*S*×1
Softmax-1		*S*×*S*×1
Tile		*S*×*S*×*D*
Input×Tile		*S*×*S*×*D*
GlobalAvegraePooling		1×1×*D*
Ветка для вычисления значений ворот		
FC-1		1×1×256
FC-2		1×1×2
Tanh-1		1×1×2
Softmax-2		1×1×2
Ветка для вычисления входа		
FC-1		1×1×256
FC-2		1×1×2
FC-2×Softmax-2		1×1×2

3.3 Выравнивание и аугментация изображений

Во всех проведенных экспериментах изображения лиц были предварительно выровнены по положению глаз и обрезаны до размера 192×192 пикселя (соотношение между размером лица и шириной «кропа» определяется коэффициентом α). Для расширения обучающей выборки спуфинга применялся ряд процедур аугментации, имитирующих различный эффекты: размытие гауссовским фильтром с размером из диапазона [11; 21], изменение разрешения с коэффициентом из диапазона [0,1; 0,2] и последующим восстановлением до исходного, добавление шума с нормальным распределением (μ=25, σ=25).

4. Результаты экспериментов

В этом разделе приводятся результаты экспериментального анализа предложенного способа детектирования спуфинг-атак на наборе данных CASIA-FASD [22]. В записи этой базы принимали участие 50 человек, для каждого из которых были сделаны 12 коротких видеороликов (3 реальных и 9 поддельных). Спуфинг-атаки включают демонстрацию фотографии на бумаге и планшете. Обучающее подмножество включает 240 записей (20 человек), тестовое – 360 записей (30 человек). В соответствии с протоколом [1] качество моделей оценивается с помощью метрики HTER=0,5×(FAR+FRR), при этом классификация видео осуществляется по 30 случайно выбранным кадрам.

Эксперименты были проведены на фреймворке TensorFlow: оптимизатор – Adadelta; начальная скорость обучения – 0,3; уменьшение скорости – экспоненциальное с шагом 500 и коэффициентом 0,01; размер партии – 32; количество итераций – 20000.

В таблице 4 приведены результаты тестирования предложенной модели CHC с одним входом. Классификация изображений при α = 1 является наиболее сложной задачей, т.к. доступна информация только о текстуре лицевой области. Лучшее качество достигается при α = 3, когда границы фотографии и рамки планшета попадают в область интереса. При этом включение модулей внимания приводит к повышению HTER только в этом случае, что подтверждает высокую информативность контекста вокруг лица для рассматриваемой задачи. В тоже время, даже если изображение лица занимает весь угол обзора камеры (рис. 1), спуфинг-атака по-прежнему может быть идентифицирована. Дальнейшие эксперименты были проведенны с α = 3.

Таблица 4. Тестирование на базе CASIA-FASD (HTER, %)

α	1	1,5	2	3
				
off	12,34	11,48	9,82	4,26
on	12,96	12,41	9,82	3,33

***** off/on attention mechanism

В таблице 5 приведены результаты тестирования алгоритмов предобработки исходных изображений и их комбинаций. В соответствии с полученными данными можно сделать следующие выводы:

1) Среди различных вариантов цветовых пространств лучшим является HSV. Это объясняется свойством данного пространства, в котором тон и яркость пикселей разделяются на разные каналы, поэтому оно часто используется для выделения участков кожи.

The publication of the conference proceedings was supported by the RFBR, grant No. 18-07-20045\18

2) Популярные LBP-дескрипторы плохо поддаются классификации сверточной нейронной сетью.
3) Лучшими по соотношению качество/производительность являются вельветы Хаара.
4) Наиболее высокое качество для СНС с одним входом достигается при применении нормализации Tan-Triggs, которое устраняет вариации в освещенности и усиливает градиенты на границах. Однако вычислительная сложность такого преобразования значительно выше, чем у остальных.
5) Наилучшим решением является сочетание нормализации Tan-Triggs с вельветами Хаара. При этом дальнейшее увеличение числа входов не приводит к уменьшению HTER.

Таблица 5. Сравнение эффективности различных алгоритмов предобработки на базе CASIA-FASD (α = 3)

Преобразование	HTER. %
RGB	8,89
HSV	7,96
YCbCr	13,33
LBP	10,19
Haar	6,3
Tan-Triggs	1,85
Tan-Triggs, Haar	**1,67**
Tan-Triggs, D4	3,15
Tan-Triggs, LBP	5,0
Tan-Triggs, LBP, Haar	4,82
Tan-Triggs, RGB, LBP, Haar	7,59

Сравнение предложенного подхода к решению задачи обнаружения витальности лица с другими современными алгоритмами, основанными на нейронных сетях, приведено в таблице 6. Разработанная мультимодальная сверточная нейронная сеть с механизмом внимания в сочетании с нормализацией Tan-Triggs и вельветами Хаара достигает state-of-the-art результатов на бенчмарке CASIA-FASD, превосходя другие алгоритмы.

Таблица 6. Сравнение нейросетевых методов обнаружения витальности лица на базе CASIA-FASD

Алгоритм	EER, %	HTER, %
Fine-tuned VGG-Face [9]	5,20	–
DPCNN [9]	4,50	–
Yang et al. [20]	4,92	–
LSTM-CNN [19]	5,17	5,93
Patch-based CNN [1]	4,44	3,78
Depth-based CNN [1]	2,85	2,52
Fusion [1]	2,67	2,27
Волкова и др.[23]	–	8,71
Предложенный	–	**1,67**

5. Заключение

В данной работе предложен эффективный способ обнаружения спуфинг-атак на системы лицевой биометрии по одной фотографии, превосходящий качество современных нейросетевых алгоритмов на бенчмарке CASIA-FASD. Однако более сложные варианты спуфинга, основанные на 3D-реконструкции лица, не были рассмотрены. Идентификация подложной 3D-маски высокого качества явятся трудной задачей и, возможно, может быть решена только с использованием видеопоследовательности, поэтому предложенный подход требует дальнейшего исследования.

Исследование проводится в рамках соглашения о предоставлении субсидии № 14.578.21.0189 (RFMEFI57816X0189)

6. Литература

[1] Atoum, Y. Face anti-spoofing using patch and depth-based CNNs, Biometrics, 2017
[2] Bharadwaj, S. Computationally efficient face spoofing detection with motion magnification, CVPRW, 2013
[3] Boulkenafet, Z. Face anti spoofing using speeded up robust features and fisher vector encoding, IEEE Signal Processing Letters, vol. 24, issue: 2, 2017
[4] Erdogmus, N. Spoofing in 2D face recognition with 3D masks and anti-spoofing with Kinect, in Proc. IEEE Biometrics, Theory, 2013, pp. 1-6
[5] Kim, G. Face liveness detection based on texture and frequency analyses. In International Conference on Biometrics, 2012, pp. 67–72
[6] Kim, I. Face spoofing detection with highlight removal effect and distortions. systems, SMC, 2016
[7] Kim, S. Face Liveness Detection Using Defocus, Sensors, vol. 15, 2015, pp. 1537–1563
[8] Li, J. Live face detection based on the analysis of fourier spectra, Biometric technology for human identification, Proc. SPIE, vol. 5404, 2004, pp. 296-303
[9] Li, L. An original face anti-spoofing approach using partial convolutional neural network. In IPTA, 2016.
[10] Liu Y. Learning deep models for face anti-spoofing: binary or auxiliary supervision, CVPR, 2018
[11] Mahmood, Z. A review on state-of-the-art face recognition approaches.. Fractals, 2017
[12] Mohan, K. Object face liveness detection with combined HOGlocal phase quantization using fuzzy based SVM classifier, 2017
[13] Patel, K. Live face video vs. spoof face video: use of moire patterns to detect replay video attacks. ICB, 2015, pp. 8-15
[14] Rodriguez, P. A painless attention mechanism for convolutional neural networks, ICLR, 2018, pp. 1-8
[15] Souza, L. How far did we get in face spoofing detection?. arxiv:1710.09868, 2017, pp. 1-48
[16] Tan X. and Triggs B., Enhanced local texture feature sets for face recognition under difficult lighting conditions, IEEE Trans. Image Process, 2010, pp. 1635–1650
[17] Wang S.-Y. Face liveness detection based on skin blood flow analysis, 2017
[18] Wang, T. Face liveness detection using 3D structure recovered from a single camera, international conference on biometrics, 2013.
[19] Xu, Z. Learning temporal features using LSTM-CNN architecture for face anti-spoofing. ACPR, 2015, pp. 41-45
[20] Yang, J. Learn convolutional neural network for face anti-spoofing. arXiv preprint arXiv:1408.5601, 2014
[21] Yuan, X. Adversarial examples: attacks and defenses for deep learning, arXiv:1712.07107, 2017, pp. 1-22
[22] Zhang, Z. A face antispoofing database with diverse attacks. Proceedings of 5th IAPR International Conference on Biometrics, IEEE (2012), pp. 26-31
[23] Волкова, С. Применение сверточных нейронных сетей для решения задачи противодействия атаке спуфинга в системах лицевой биометрии / С. Волкова, Ю. Н. Матвеев // Научно-технический вестник информационных технологий, механики и оптики, Т.17, № 4, с. 702-710, 2017.

Об авторах

Калиновский Илья Андреевич, к.т.н., научный сотрудник ООО «ЦРТ-инновации». Его e-mail kalinovskiy@speechpro.com.

Лаврентьева Галина Михайловна, аспирантка каф. РИС Университета ИТМО, научный сотрудник ООО «ЦРТ-инновации». Ее e-mail lavrentyeva@speechpro.com.

Работа опубликована при финансовой поддержке РФФИ, грант 18-07-20045\18