АНАЛИЗ И СИНТЕЗ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 004.855.5

ИСПОЛЬЗОВАНИЕ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ОБНАРУЖЕНИЯ ПОДМЕНЫ ЛИЦА ЕГО ИЗОБРАЖЕНИЕМ

© **Д. В. Пакулич**[1,2], **С. А. Алямкин**[2]

[1] *Новосибирский государственный университет, 630090, г. Новосибирск, ул. Пирогова, 2*
[2] *ООО «Экспасофт», 630090, г. Новосибирск, ул. Николаева, 12*
E-mail: d.pakulich@expasoft.tech

Увеличение доли компьютерного зрения для систем биометрической идентификации и использование его в качестве меры безопасности приводит к росту попыток фальсификации. В связи с этим увеличивается и количество способов автоматического определения подобных ситуаций. Но как и большинство систем, использующих компьютерное зрение в различных ситуациях, точность распознавания атак может снижаться в некоторых случаях. В предлагаемой работе рассмотрены существующие подходы к распознаванию спуфинг-атак и дана оценка их устойчивости к изменению условий записи данных.

Ключевые слова: свёрточные нейронные сети, распознавание спуфинг-атак, глубокие нейронные сети, компьютерное зрение.

DOI: 10.15372/AUT20210411

Введение. В настоящее время существует множество различных программ для работы с биометрическими данными человека: распознавание возраста человека по лицу и даже по рукам, определение его настроения и другие программы. В первую очередь биометрические данные, особенно лицо человека, используются в целях идентификации, что позволяет повысить безопасность и обеспечить удобный доступ человеку к закрытым объектам, где установлены устройства к данным различных устройств. Однако, где осуществляется безопасность, там появляются и способы её нарушить. В представленной работе рассматривается один из таких способов — спуфинг-атака (spoofing attack (подмена)).

В контексте сетевой безопасности это — ситуация, в которой один человек или программа успешно маскируется под другую путём фальсификации данных и позволяет получить незаконные преимущества по определению. К ним можно отнести доступ к закрытым частям сайта с помощью изменения запроса, мошенничество с помощью звонков по телефону в целях позиционирования себя как другого человека. С развитием компьютерного зрения спуфинг-атаки стали применяться в этой области.

Спуфинг-атаки системы распознавания лиц делятся на три основных типа: воспроизведение видеозаписи, распечатанная фотография или изображение на экране и использование маски. Эти типы атак сильно различаются, поэтому чаще всего на каждый из них разрабатывается отдельное решение. Развитие нейронных сетей привело к появлению универсальных решений, которые могут распознавать все представленные им атаки. Помимо универсальности эти методы показывают одни из лучших результатов в настоящее время.

Также важной частью выбора метода для распознавания атак является входной формат данных. Видеозаписи с попыткой доступа более информативны, чем один кадр. Много

камер, кроме самой картинки, могут также получать и карты глубины или оптический поток. Наконец, важным фактором является качество самой камеры, позволяющее получить больше деталей, которые помогают в решении поставленной задачи.

Целью данной работы является определение эффективности существующих подходов по распознаванию спуфинг-атак с помощью нейронных сетей на одном отдельном кадре.

Одни из самых первых способов распознавания атак основываются на детектировании моргания. К примеру, можно распознавать человека с открытыми и закрытыми глазами и следить за изменением состояния в последовательности кадров. Такой подход до сих пор является одним из самых популярных.

В [1] представлена одна из первых баз данных REPLAY-ATTACK и предложен один из простых вариантов для детектирования атак, основанный на подсчёте локальных бинарных паттернов (LBP, local binary pattern) для различных зон лица после выделения его из кадра. Самый простой шаблон LBP для конкретного пикселя, обычно обозначаемый как LBP размером 3×3, формируется путём сравнения значений интенсивности этого пикселя со значениями интенсивности пикселей в его окрестности 3×3. Таким образом, каждому пикселю присваивается метка со значением от 0 до 255 (2^8–1). В случае однородного LBP (LBPu2) рассматриваются только те метки, которые содержат не более двух переходов: 0 – 1 или 1 – 0. Вектор признаков изображения или область изображения образуется путём вычисления гистограммы меток пикселей. Для каждой выделенной области лица строится гистограмма, на основе которой с помощью обученного SVM-классификатора предсказывается, является ли данное лицо результатом атаки или реальным человеком.

В [2] предложено вместо RGB-разбиения на цвета использовать HSV и YCbCr для подсчёта локальных бинарных паттернов. После подсчёта паттернов, как и в предыдущем методе, строились гистограммы и обучался SVM-классификатор для распознавания атак. В результате их исследования комбинация HSV и YCbCr показала значительное повышение точности, что позволило ей превзойти существовавшие на тот момент решения.

В некоторых работах находят артефакты, возникающие при записи экрана на камеру (муар). В [3] предложено искать артефакты в видеозаписи периодического узора на изображениях, вызванного их наложением.

Методы, описанные в [4–7], обучают глубокую нейронную сеть классификации между реальными людьми и спуфингом. В [8–10] предлагаются дополнительные данные, такие как карта глубины лица и сигнал фотоплетизмографии (регистрация кровяного потока с использованием источника инфракрасного или светового излучения), чтобы помочь сети изучить более универсальные функции.

Помимо детектирования спуфинга при распознавании лиц, методы машинного обучения используются для детектирования спуфинга аудиозаписей [11, 12].

Базы данных. Для работы были выбраны три базы данных: Replay-Attack, SiW, ROSE. Каждая база представляет собой наборы видеозаписей реальных людей и спуфинг-атак, собранных в разных условиях с помощью различных устройств. Условия сбора данных варьируются внутри баз и в то же время различаются между базами (например, нет повторяющихся устройств).

База данных Replay-Attack [1] состоит из видеозаписей реальных людей и попыток атак 50 клиентов. Для каждого человека было записано несколько видеороликов в различных условиях освещённости и были сделаны фотографии с высоким разрешением. В данных представлены три типа атак: с использованием напечатанной фотографии, фотографии на телефоне и с помощью видеографии на телефоне. В соответствии с условиями крепления устройств с поддельными лицами перед камерой атаки были разделены на атаки с ручной поддержкой (устройства для атаки держались оператором) и атаки с фиксированной поддержкой (устройства для атаки были установлены на фиксированном уровне). Для корректной оценки точности общий набор видеороликов делится на три подмножества, в

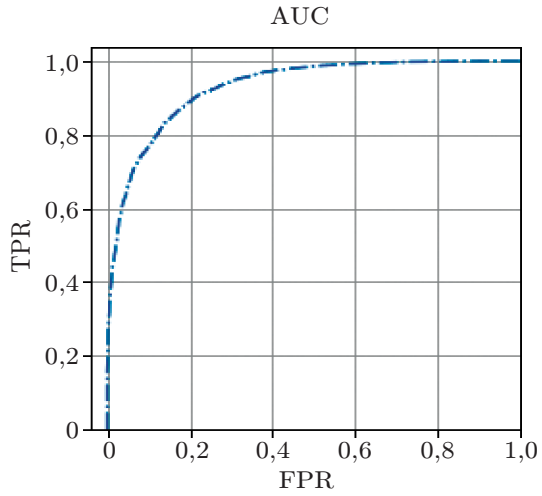

Рис. 1. График ROC-кривой с AUC, равной 0,929

которых нет пересекающихся лиц, чтобы исключить участие объектов для тестирования в обучении моделей.

База данных SiW (Spoof in the Wild) [13] представляет собой реальные видеозаписи без атак и видеозаписи атак на 165 человек. Для каждого объекта есть 8 реальных записей и до 20 фальшивых — всего 4478 видеозаписей. Все видеозаписи имеют скорость 30 кадр/с, продолжительность около 15 с и разрешение 1080P HD. Реальные записи собраны в четыре сеанса с вариациями расстояния, позы, освещения и выражений лиц. Поддельные видеозаписи собираются с помощью нескольких видов атак, таких как напечатанные на бумаге и воспроизведённые на устройстве.

База данных ROSE-Youtu Face Liveness Detection [14] охватывает широкий спектр условий освещения, моделей камер, типов атак и состоит из 4225 видеозаписей для 25 человек (доступны 3350 для 20 человек).

Метрики. Так как модели предсказывают вероятность того, что изображение относится к определённому классу (реальное лицо или спуфинг), то в зависимости от выбора порога меняются результаты классификации. В связи с этим существуют различные метрики для оценки результата.

Для подсчёта классификации обычно находят долю правильно и неправильно определённых объектов позитивного и негативного классов (True Positive Rate (TPR), True Negative Rate (TNR), False Positive Rate (FPR), False Negative Rate (FNR)). В задаче детектирования спуфинга часто применяют метрики EER (Equal Error Rate) и HTER (Half Total Error Rate). Для нахождения EER используют валидационную выборку. На ней определяется порог вероятности, при котором значения FPR и FNR равны. Значение FPR при этом пороге называется EER. Метрика HTER вычисляется на тестовой выборке и равна полусумме FPR и FNR при пороге, полученном при вычислении EER.

Однако эти метрики зависят от данных, на которых вычислялась EER, поскольку порог может сильно варьироваться. Соответственно для экспериментов в данной работе использовалась AUC или площадь под кривой, показывающая отношение TPR к FPR (рис. 1).

ROC-кривая (receiver operating characteristic) — график, позволяющий оценить качество бинарной классификации и отображающий соотношение долей объектов от общего количества носителей признака, верно классифицированных как несущие признак (TPR), и долей объектов от общего количества объектов, не несущих признака, ошибочно классифицированных как несущие признак (FPR) при варьировании порога решающего правила.

Количественную интерпретацию ROC даёт показатель AUC (area under ROC curve, площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций (см. рис. 1). Чем ближе к единице показатель AUC, тем качественнее классификатор, при этом значение 0,5 демонстрирует непригодность выбранного метода классификации (соответствует случайному гаданию). Значения ниже 0,5 указывают на обратную связь между предсказанием классификатора и реальным значением, т. е. при значении показателя AUC, равном 0, каждое предсказание будет неверным, однако если поменять предсказанные классы бинарной классификации, то каждое предсказание станет верным, что соответствует значению AUC, равному 1.

Сравнение существующих подходов.

Архитектура нейронных сетей. Рассмотрим три модели, представленные в [15, 5, 16].

В [15] предлагается подход, основанный на немного модифицированной архитектуре AlexNet. На вход сети подаётся изображение размером 224×224 пикселя. Выходом из сети является вектор из двух элементов (принадлежность изображения к классам): реальная фотография или спуфинг. В качестве функции потерь используется перекрёстная энтропия.

В [5] рассматривается архитектура, основанная на архитектуре VGG. На вход подаётся также изображение размером 224×224 пикселя, выходом из сети является вектор из двух элементов. Берётся предобученная сеть для распознавания лиц и заменяется последний полносвязный слой с изменением при этом классификации 2622 класса (количество людей, использованных для распознавания лиц) всего на 2 класса. После сеть дообучается на данных по распознаванию.

Предложенная в [16] нейронная сеть основана на более современной архитектуре DenseNet. Основная идея DenseNet состоит в том, чтобы соединить каждый слой со всеми остальными слоями (с тем же размером карты признаков) прямой связью. Для каждого слоя в качестве входных данных используются карты признаков из предыдущих слоёв. Эта реализация уменьшает проблему исчезающего градиента, поскольку блоки нейронной сети добавляют короткие пути от входов к выходам. В каждом слое карты признаков получаются путём объединения предыдущих карт признаков.

На вход подаётся также изображение размером 224×224 пикселя. Выходом из нейронной сети помимо вектора из двух элементов являются две карты признаков размером 14×14 элементов, цель которых — выявление признаков спуфинг-атаки отдельно на каждой из 196 областей.

Проведение эксперимента. В ходе эксперимента каждая из баз данных: Replay-Attack,

Таблица 1

Сравнение результатов тестирования моделей, обученных на разных базах данных

Модель	База данных для обучения	База данных для тестирования		
		Replay-Attack	SiW	ROSE
AlexNet	Replay-Attack	0,952	0,546	0,558
	SiW	0,628	0,993	0,615
	ROSE	0,534	0,663	0,904
VGG	Replay-Attack	0,958	0,733	0,516
	SiW	0,613	0,999	0,792
	ROSE	0,638	0,791	0,891
DenseNet	Replay-Attack	1,000	0,751	0,618
	SiW	0,668	1,000	0,797
	ROSE	0,822	0,929	0,992

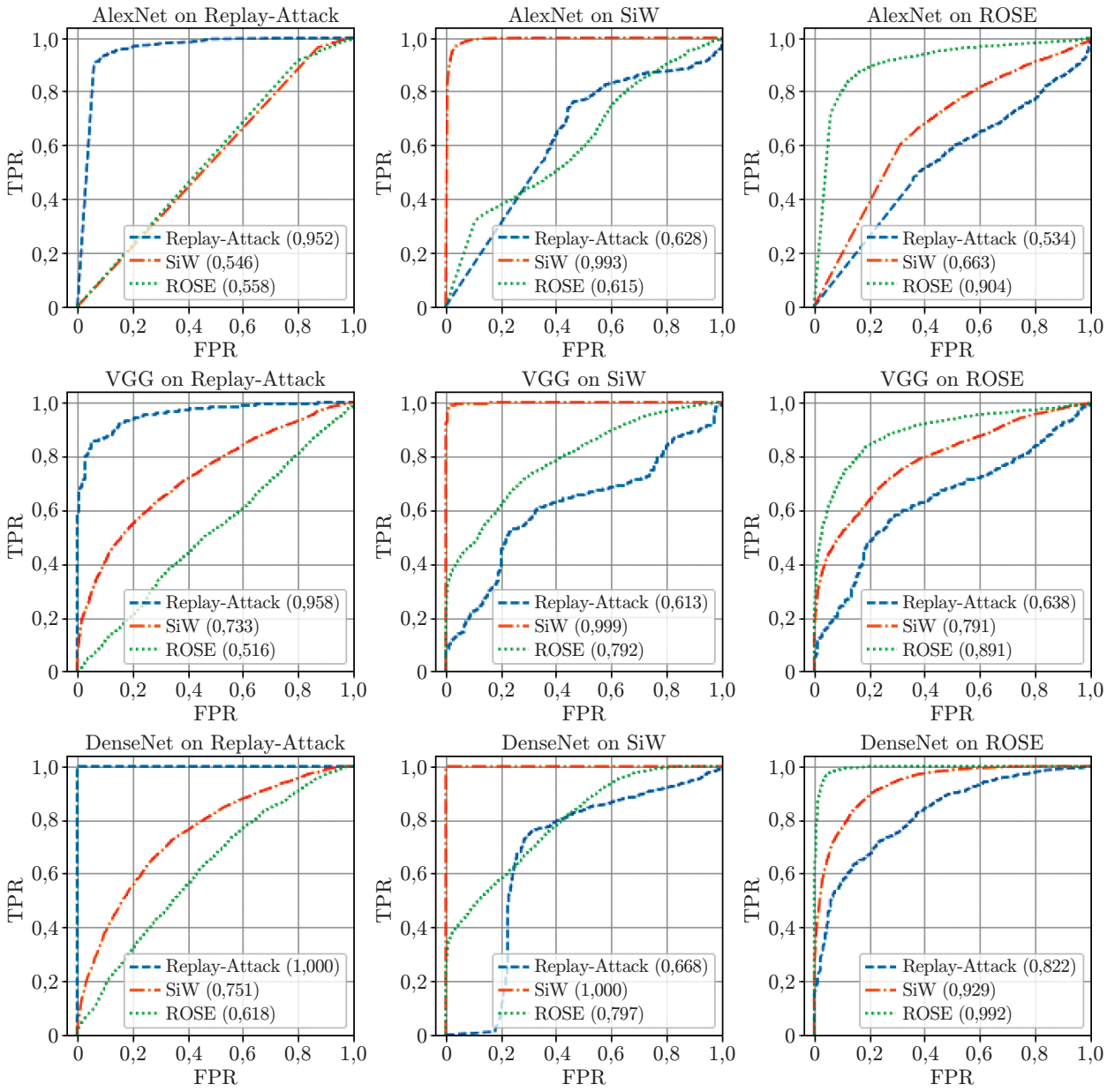

Рис. 2. Сравнение результатов тестирования моделей, обученных на разных базах данных. В скобках указаны значения показателя AUC для тестирования на соответствующей базе данных

SiW и ROSE — была разделена на три выборки с непересекающимися лицами (для обучения, валидации и тестирования).

Из видеозаписей были получены пять кадров с шагом 0,1 с. На всех кадрах производился поиск лица с помощью специально обученной нейронной сети, затем оно вырезалось. Каждое лицо приводилось к размеру 224×224 пикселя.

Затем на специальных частях были обучены три рассмотренные сети. На каждой из баз сети обучались раздельно. Для контроля над переобучением использовались их валидационные части. В результате было получено девять моделей.

На каждой из моделей проводилось тестирование тестовой части всех трёх баз данных. Результаты тестирования представлены в табл. 1 и на рис. 2. Из результатов экс-

перимента видно, что значение показателя AUC значительно меньше на базах данных, не участвующих в обучении.

Из полученных результатов можно сделать вывод, что модели, хорошо работающие на тестовых выборках баз данных, где они обучались, могут показывать значительно худший результат в других базах данных, так как каждая база данных, несмотря на кажущуюся вариативность, обладает рядом признаков, характерных именно для неё. Чаще всего причиной этому являются особенности техники, производящей сбор данных, или условия сбора данных, которые остаются похожими для различных выборок, но сильно отличаются у разных баз данных.

Решением этой проблемы в компьютерном зрении занимается доменная адаптация. Суть её состоит в адаптации модели, обученной на одном домене-источнике, к условиям, характерным для целевого домена. В задаче по спуфингу это могут быть артефакты, специфичные для устройств, с помощью которых собираются данные, ракурс съёмки, характерное освещение или задний фон.

Заключение. С развитием компьютерного зрения появляется проблема фальсификации. В связи с этим возникает необходимость в автоматическом распознавании таких случаев. Существует множество различных подходов, и большинство из них использует машинное обучение в общем и нейронные сети в частности. Для обучения и тестирования собраны различные базы данных, содержащие фото- и видеопримеры реальных людей и атак.

Выявлено, что подходы, использующие машинное обучение, показывают хороший результат тестирования на данных предложенных баз, на которых они обучались, но при тестировании на других данных точность резко падает. Это возникает в первую очередь из-за различий устройств, с помощью которых делаются снимки лиц, разницы в освещении помещений и других условий сбора данных. Данная проблема является общей для всего компьютерного зрения, но в такой задаче она проявляется наиболее сильно. Для решения проблемы предлагаются различные методики, которые позволяют частично нивелировать различия данных.

В дальнейшем планируется разработать универсальный метод, точность которого не будет сильно различаться при тестировании на любых данных. В этом может помочь выявление артефактов, специфичных для устройств, и адаптация моделей именно под них. Также можно установить не только различие между реальным лицом и подменой, но и сам факт попытки спуфинга. Для этого необходимо определить устройство, с помощью которого злоумышленник пытается осуществить спуфинг.

СПИСОК ЛИТЕРАТУРЫ

1. **Chingovska I., Anjos A., Marcel S.** On the effectiveness of local binary patterns in face antispoofing // Proc. of the Intern. Conference on Biometrics Special Interest Group (BIOSIG 2012). Darmstadt, Germany, 6–7 Sept., 2012. P. 1–7.
2. **Boulkenafet Z., Komulainen J., Hadid A.** Face anti-spoofing based on color texture analysis // Proc. of the IEEE Intern. Conference on Image Processing (ICIP 2015). Quebec City, Canada, 27–30 Sept., 2015. P. 2636–2640.
3. **Patel K., Han H., Jain A. K., Ott G.** Live face video vs. spoof face video: Use of Moiré patterns to detect replay video attacks // Proc. of the Intern. Conference on Biometrics (ICB 2015). Phuket, Thailand, 19–22 May, 2015. P. 98–105.
4. **Feng L., Po L. M., Li Y. et al.** Integration of image quality and motion cues for face anti-spoofing: A neural network approach // Journ. Visual Communication and Image Representation. 2016. **38**. P. 451–460.

5. **Li L., Feng X., Boulkenafet Z. et al.** An original face anti-spoofing approach using partial convolutional neural network // Proc. of the 6th Intern. Conference on Image Processing Theory, Tools and Applications (IPTA 2016). Oulu, Finland, 12–15 Dec., 2016. P. 1–6.
6. **Patel K., Han H., Jain A. K.** Cross-database face antispoofing with robust feature representation // Proc. of the Conference on Chinese Conference on Biometric Recognition (CCBR 2016). Chengdu, China, 14–16 Oct., 2016. P. 611–619.
7. **Yang J., Lei Z., Li S. Z.** Learn convolutional neural network for face anti-spoofing // Comp. Vis. Pattern Recogn. Cornell Univers., 2014. URL: https://arxiv.org/abs/1408.5601 (дата обращения: 13.04.2021).
8. **Atoum Y., Liu Y., Jourabloo A., Liu X.** Face anti-spoofing using patch and depth-based CNNs // Proc. of the Conference on IEEE Intern. Joint Conference on Biometrics (IJCB). Denver, USA, 1–4 Oct., 2017. P. 319–328.
9. **Shao R., Lan X., Li J., Yuen P. C.** Multi-adversarial discriminative deep domain generalization for face presentation attack detection // Proc. of the Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 15–20 June, 2019. P. 10015–10023.
10. **Yang X., Luo W., Bao L. et al.** Face anti-spoofing: Model matters, so does data // Proc. of the Conference on Computer Vision and Pattern Recognition. Long Beach, USA, 15–20 June, 2019. P. 3502–3511.
11. **Tak H., Patino J., Nautsch A. et al.** Spoofing attack detection using the non-linear fusion of sub-band classifiers // Comp. Vis. Pattern Recogn. Cornell Univers., 2020. URL: https://arxiv.org/abs/2005.10393 (дата обращения: 13.04.2021).
12. **Dinkel H., Chen N., Qian Y., Yu K.** End-to-end spoofing detection with raw waveform CLDNNS // Proc. of the IEEE Intern. Conference on Acoustics, Speech and Signal Processing (ICASSP 2017). New Orleans, USA, 5–9 March, 2017. P. 4860–4864.
13. **Liu Y., Jourabloo A., Liu X.** Learning deep models for face anti-spoofing: Binary or auxiliary supervision // Proc. of the Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 18-22 June, 2018. P. 389–398.
14. **Li H., Li W., Cao H. et al.** Unsupervised domain adaptation for face anti-spoofing // IEEE Transactions on Information Forensics and Security. 2018. **13**, N 7. P. 1794–1809.
15. **Волкова С. С., Матвеев Ю. Н.** Применение свёрточных нейронных сетей для решения задачи противодействия атаке спуфинга в системах лицевой биометрии // Науч.-техн. вестн. информационных технологий, механики и оптики. 2017. **17**, № 4. С. 702–710.
16. **George A., Marcel S.** Deep pixel-wise binary supervision for face presentation attack detection // Proc. of the Intern. Conference on Biometrics (ICB). Crete, Greece, 4–7 June, 2019. P. 1–8.

Поступила в редакцию 12.04.2021
После доработки 13.05.2021
Принята к публикации 17.05.2021