Метод обнаружения технологии манипуляции с идентичностью лиц с применением сверточных нейронных сетей *

С. С. Волкова

Вологодский государственный университет, Вологда, Россия

Аннотация. В работе предложен метод противодействия атакам спуфинга, позволяющий повысить устойчивость систем биометрической аутентификации по изображению лица к атакам спуфинга на модуль ввода биометрической информации цифровыми средствами манипуляции с идентичностью лиц. Метод обнаружения манипуляций с подменой лиц цифровыми средствами (DeepFake) основывается на сверточной нейронной сети, обученной на большом наборе данных, содержащих различные типы манипуляций, изображения разного качества и большое число идентичностей, что позволило получить точность не менее 99%. Результаты экспериментов также показывают эффективность предложенного подхода при сравнении его с другими известными методами, протестированными на том же наборе данных. Предложенный метод может использоваться для повышения защищенности систем биометрической аутентификации путем снижения риска несанкционированного доступа.

Ключевые слова: биометрическая аутентификация, спуфинг, подмена лиц, сверточная нейронная сеть, распознавание.

DOI 10.14357/20718594220206

Введение

За последние 5 лет число киберпреступлений выросло в 11 раз. По данным управления правовой статистики и информационных технологий Генпрокуратуры в 2021 году на преступления, совершенные с использованием информационно-телекоммуникационных технологий или в сфере компьютерной информации, приходится одно из четырех регистрируемых преступлений. При этом в последние годы лидерство занимают преступления, связанные с дистанционным банковским обслуживанием [1]. Поэтому разработка и развитие систем биометрической аутентификации, когда аутентификационная информация - неотъемлемая часть пользователя, крайне актуальна, что подтверждается темпами развития рынка биометрических технологий, использующихся в банковском секторе для подтверждения личности с использованием фотографии или голоса клиентов, а также с целью подтверждения переводов крупных сумм с мобильных устройств.

Однако системы биометрической аутентификации подвержены атакам спуфинга, когда неавторизованные клиенты (злоумышленники) могут попытаться обмануть систему путем подмены «живого» изображения лица/записи голоса авторизованного клиента его фотографией или видео-, аудио - файлом. Особенно остро стоит вопрос с обнаружением манипуляций с подменой лиц цифровыми средствами.

*Работа выполнена при финансовой поддержке Правительства Вологодской области в рамках государственного научного гранта.

✉ *Волкова Светлана Сергеевна. E-mail: malysheva.svetlana.s@gmail.com*

Методика синтеза изображений DeepFake (англ. Deep Learning – «глубокое обучение», Fake - подделка) в последние годы активно развивается и человеческий глаз не всегда способен отличить подделку от оригинального изображения.

Таким образом, устойчивость биометрической системы аутентификации пользователей информационных систем к атакам спуфинга на модуль ввода биометрической информации цифровыми средствами манипуляции с идентичностью лиц является одной из важнейших задач для предотвращения несанкционированного доступа к информационной системе, предотвращению киберпреступлений, а также противодействия кибертерроризму.

Существуют различные техники манипуляции с лицами цифровыми средствами. Одни из них направлены на манипуляции с выражением лиц (например, Face2Face) [2], другие — на подмену идентичности лиц, когда лицо одного человека подменяется лицом другого [3, 4]. Вторые методы наиболее часто используются для поиска уязвимостей систем биометрической аутентификации.

Целью работы является разработка метода повышения устойчивости систем биометрической аутентификации пользователей информационных систем по изображению лица к атакам спуфинга цифровыми средствами манипуляции с изображениями лиц. Метод для обнаружения попытки применения способов манипуляции, основывается на анализе текстуры и решение о том, является ли представленный образец поддельным, может быть принято по одному кадру. Подход базируется на сверточной нейронной сети для извлечения и классификации вектора лицевых признаков.

Статья организована следующим образом: в первом разделе обобщается современное состояние проблемы обнаружения применения технологии манипуляции с идентичностью лиц. Второй раздел описывает предлагаемый метод, в третьем – представлены результаты экспериментов. Четвертый раздел посвящен применению разработанного метода обнаружения манипуляций с идентичностью лиц для повышения защищенности биометрических систем. В заключении подводится итог всему исследованию, а также предлагаются направления дальнейшего развития.

1. Современные методы обнаружения технологии манипуляций с идентичностью лиц

Методы обнаружения технологии манипуляции с идентичностью лиц можно условно разделить на две категории: методы, направленные на поиск артефактов и методы, направленные на построение универсальных классификаторов.

Методы, направленные на поиск артефактов, пытаются выделить особенности, которые генерируются самими средствами цифровой манипуляции [5-9], поскольку способы манипуляции с идентичностью лиц могут оставлять дефекты и искажения, которые могут остаться незаметными для человеческого глаза. Например, генеративно-состязательные сети могут оставлять текстурные отпечатки.

Методы, направленные на построение глобального классификатора [10-20], не фокусируются на отдельных особенностях, оставляемых средствами манипуляции. Именно методы второй категории активно развиваются в последнее время, поскольку в связи с развитием технологий манипуляции с лицами, когда человеческий глаз не всегда способен отличить была ли применена манипуляция или нет, а сами способы подмены практически не оставляют видимых дефектов и искажений, методы обнаружения, основанные на артефактах, не дают должного качества распознавания.

Методы, основанные на построении универсального классификатора, все чаще используют технологии глубокого обучения либо для построения глобального классификатора [10-17, 21], либо для поиска выбросов в данных [18-20], что позволяет получить информацию, что данные не соответствуют ожидаемому распределению. При этом методы могут основываться как на одном изображении [10, 11, 21], так и на группе кадров [16, 17].

В качестве архитектур нейронных сетей использовались XceptionNet [10], EfficientNet [17], MesoInception-4 — модель с небольшим числом слоев для того, чтобы сосредоточится на мезоскопических свойствах изображений [11]. Для поиска локальных особенностей в [22] была предложена облегченная модель, состоящая из 10 слоев. Также глобальные классификаторы были представлены в работах [12-15].

Несмотря на то, что существует множество различных подходов для решения задачи обнаружения попытки манипуляции с идентичностью лиц, все они не всегда показывают достаточное качество распознавания.

2. Предлагаемый метод

Для обнаружения ситуации, когда злоумышленник маскируется под авторизованного клиента, используя цифровые методы манипуляции с идентичностью лиц, предлагается метод, основывающийся на классификации признаков, полученных сверточной нейронной сетью. Схема работы метода определения подмены лиц цифровыми средствами представлена на Рис. 1 и состоит из двух последовательных блоков: блока предварительной обработки изображения и блока извлечения и классификации признаков лица.

Блок предварительной обработки изображения включает в себя процедуру обнаружения лица и контрольных точек и процесс его выравнивания. Для обнаружения лица используется метод, описанный в [23], для поиска контрольных точек — в [24].

Для выравнивания лица используется четыре контрольные точки: внешние углы глаз и внешние углы губ. Необходимость использования процедуры выравнивания изображения лица обусловлена неустойчивостью нейросетевых алгоритмов к афинным преобразованиям изображения (поворот и масштаб).

Процедура выравнивания выполняется следующим образом:

1. Поворот лица в горизонтальной плоскости, используя координаты внешних углов глаз, что компенсирует наклон головы вправо и влево, вращение вокруг оси через нос и затылок. Угол поворота определяется по формуле:

$$\alpha = \text{atan}(\frac{y_{E_L} - y_{E_R}}{x_{E_L} - x_{E_R}}), \quad (1)$$

где (x_{E_L}, y_{E_L}) — координата внешнего угла левого глаза; (x_{E_R}, y_{E_R}) — координата внешнего угла правого глаза.

2. Масштабирование повернутого изображения до размеров, когда расстояние

$$L = \frac{(y_{M_L} - y_{E_L}) + (y_{M_R} - y_{E_R})}{2},$$

где (x_{E_L}, y_{E_L}) — координата внешнего угла левого глаза; (x_{E_R}, y_{E_R}) — координата внешнего угла правого глаза; (x_{M_L}, y_{M_L}) — координата левого угла губы; (x_{M_R}, y_{M_R}) — координата правого угла губы, не превышает заданную величину.

3. Вычисление координат вершин прямоугольника обрезки по следующим формулам: $A_0(x_{E_R} - 0.8 \cdot L; y_{E_R} - L)$, $A_1(x_{E_L} + 0.8 \cdot L; y_{E_R} - L)$, $A_2(x_{E_L} + 0.8 \cdot L, ; y_{M_L} + 1.75 \cdot L)$, $A_3(x_{E_R} - 0.8 \cdot L; y_{M_L} + 1.75 \cdot L)$.

4. Обрезка изображения по координатам полученного прямоугольника.

Результат выполнения процедур поиска и выравнивания изображения лица показан на Рис. 2.

Для выделения и классификации лицевых признаков используется сверточная нейронная сеть типа MobileNet [25], предварительно обученная на ImageNet [26]. На вход сети подается предварительно обработанное в предыдущем блоке изображение лица в RGB-формате, размера 224x224 пикселя. Каждый пиксель нормализован к диапазону [0, 1]. В Табл. 1 показаны параметры используемой архитектуры нейронной сети.

В таблице приняты следующие обозначения: Conv2d – сверточный слой; ReLU6 – функция активации, имеющая вид $f(x) = \min(\max(0, x), 6)$; BatchNorm2d — слой пакетной нормализации; AdaptiveAvgPool2d — слой пространственного объединения; Dropout — слой регуляризации, прореживания сети путем отключения нейронов сети с заданной вероятностью; Linear — полносвязный слой.

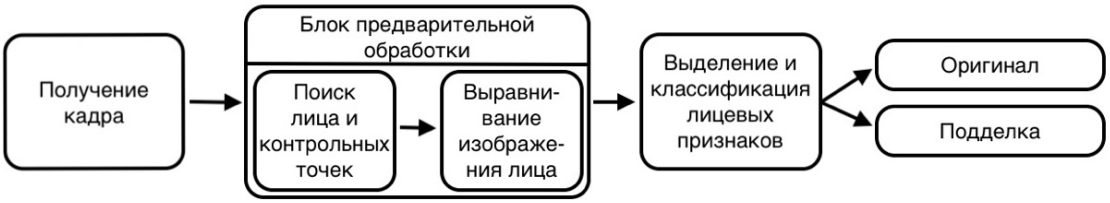

Рис. 1. Схема работы предложенного метода

(a) (b)

Рис. 2. Результат выполнения процедур поиска лица и контрольных точек (a), выравнивания изображения лица (b)

Табл. 1. Параметры архитектуры используемой нейронной сети

№	**Тип слоя/ имя**	**Количество повторений слоя**	**Размер ядра/шаг** пикс.	**Выходной размер** (ширина, пикс) × (высота, пикс) × количество
1	Input	1		[3, 224, 224]
2	Conv2d BatchNorm2d ReLU6	1	3×3/2 - -	[32, 112, 112] [32, 112, 112] [32, 112, 112]
3	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	1	3×3/1 - - 1×1/1 -	[32, 112, 112] [32, 112, 112] [16, 112, 112] [16, 112, 112] [16, 112, 112]
4	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	1	1×1/1 - - 3×3/2 - - 1×1/1 -	[96, 112, 112] [96, 112, 112] [96, 112, 112] [96, 56, 56] [96, 56, 56] [96, 56, 56] [24, 56, 56] [24, 56, 56]
5	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	1	1×1/1 - - 3×3/1 - - 1×1/1 -	[144, 56, 56] [144, 56, 56] [144, 56, 56] [144, 56, 56] [144, 56, 56] [144, 56, 56] [24, 56, 56] [24, 56, 56]
6	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	1	1×1/1 - - 3×3/2 - - 1×1/1 -	[144, 56, 56] [144, 56, 56] [144, 56, 56] [144, 28, 28] [144, 28, 28] [144, 28, 28] [32, 28, 28] [32, 28, 28]

Табл. 1. Параметры архитектуры используемой нейронной сети (продолжение)

№	Тип слоя/ имя	Количество повторений слоя	Размер ядра/шаг пикс.	Выходной размер (ширина, пикс) × (высота, пикс) × количество
7	Conv2d	2	1×1/1	[192, 28, 28]
	BatchNorm2d		-	[192, 28, 28]
	ReLU6		-	[192, 28, 28]
	Conv2d		3×3/1	[192, 28, 28]
	BatchNorm2d		-	[192, 28, 28]
	ReLU6		-	[192, 28, 28]
	Conv2d		1×1/1	[32, 28, 28]
	BatchNorm2d		-	[32, 28, 28]
8	Conv2d	1	1×1/1	[192, 28, 28]
	BatchNorm2d		-	[192, 28, 28]
	ReLU6		-	[192, 28, 28]
	Conv2d		3×3/2	[192, 14, 14]
	BatchNorm2d		-	[192, 14, 14]
	ReLU6		-	[192, 14, 14]
	Conv2d		1×1/1	[64, 14, 14]
	BatchNorm2d		-	[64, 14, 14]
9	Conv2d	3	1×1/1	[384, 14, 14]
	BatchNorm2d		-	[384, 14, 14]
	ReLU6		-	[384, 14, 14]
	Conv2d		3×3/1	[384, 14, 14]
	BatchNorm2d		-	[384, 14, 14]
	ReLU6		-	[384, 14, 14]
	Conv2d		1×1/1	[64, 14, 14]
	BatchNorm2d		-	[64, 14, 14]
10	Conv2d	1	1×1/1	[384, 14, 14]
	BatchNorm2d		-	[384, 14, 14]
	ReLU6		-	[384, 14, 14]
	Conv2d		3×3/1	[384, 14, 14]
	BatchNorm2d		-	[384, 14, 14]
	ReLU6		-	[384, 14, 14]
	Conv2d		1×1/1	[96, 14, 14]
	BatchNorm2d		-	[96, 14, 14]
11	Conv2d	2	1×1/1	[576, 14, 14]
	BatchNorm2d		-	[576, 14, 14]
	ReLU6		-	[576, 14, 14]
	Conv2d		3×3/1	[576, 14, 14]
	BatchNorm2d		-	[576, 14, 14]
	ReLU6		-	[576, 14, 14]
	Conv2d		1×1/1	[96, 14, 14]
	BatchNorm2d		-	[96, 14, 14]
12	Conv2d	1	1×1/1	[576, 14, 14]
	BatchNorm2d		-	[576, 14, 14]
	ReLU6		-	[576, 14, 14]
	Conv2d		3×3/2	[576, 7, 7]
	BatchNorm2d		-	[576, 7, 7]
	ReLU6		-	[576, 7, 7]
	Conv2d		1×1/1	[160, 7, 7]
	BatchNorm2d		-	[160, 7, 7]

Табл. 1. Параметры архитектуры используемой нейронной сети (окончание)

№	Тип слоя/ имя	Количество повторений слоя	Размер ядра/шаг пикс.	Выходной размер (ширина, пикс) × (высота, пикс) × количество
13	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	2	1×1/1 - - 3×3/1 - - 1×1/1 -	[960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [160, 7, 7] [160, 7, 7]
14	Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d ReLU6 Conv2d BatchNorm2d	1	1×1/1 - - 3×3/1 - - 1×1/1 -	[960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [960, 7, 7] [320, 7, 7] [320, 7, 7]
15	AdaptiveAvgPool2d Dropout Linear			[320, 1, 1] [320] [2]

Для обучения сети использовался PyTorch, в качестве функции потерь — перекрестная энтропия. В качестве метода стохастической оптимизации был использован MADGRAD [27] — метод, содержащий некоторые модификации метода AdaGrad [28], которые дают более высокую способность к обобщению для большого числа задач, в том числе и тех, с которыми ранее не справлялся метод Adam [29], широко использующийся для обучения нейронных сетей.

В качестве стратегии изменения скорости обучения использовался SGDR [30], при которой для каждой итерации множитель скорости обучения меняется по формуле:

$$\eta_t = \eta_{min}^{(i)} + 0.5 \cdot \left(\eta_{max}^{(i)} - \eta_{min}^{(i)}\right) \cdot \left(1 + \cos\left(\frac{\pi \cdot T_{cur}}{T_i}\right)\right), \quad (2)$$

где $\eta_{min}^{(i)}$ и $\eta_{max}^{(i)}$ — диапазон значений скорости обучения, T_{cur} фиксирует число эпох, прошедших с последнего перезапуска и обновляется на каждой итерации; $i + 1$ - й перезапуск происходит тогда, когда $T_i = T_{cur}$; T_i — задан; при каждом перезапуске значение T_{cur} обнуляется.

Начальная скорость обучения была выставлена равной 0.01, общее число эпох — 300. Для обучения использовался набор данных FaceForencisc++ [10]. Для генерации более разнообразного обучающего набора, данные были дополнены путем добавления случайного гауссова шума, размытия и сжатия JPEG, а также коррекции контраста, яркости и гаммы случайным образом.

Для обучения и тестирования модели использовался ПК со следующими характеристиками: ЦПУ Core(TM) i3 CPU 6100 @ 3.7GHz, видеокарта NVidia GeForce GTX 1080 8Гб, ОЗУ 8Гб. Графики изменения средней ошибки на обучающей и валидационной выборках представлены на Рис. 3.

3. Результаты экспериментов

Для оценки предложенного метода, а также сравнения его с другими известными способами обнаружения манипуляции с идентичностью лиц цифровыми средствами, использовался набор данных FaceForencis++ [10]. Он содержит изображения, полученные из 1000 видеороликов, все данные разделены на оригиналы и подделки, которые в свою очередь разделены на 3 категории: Face2Face (подмена выражения лиц), FaceSwap и DeepFake (подмена идентичности лиц). Особый интерес для данной работы представляют последние две категории, так как направлены на подмену идентичности лиц. Однако для оценки качества работы метода поиска манипуляций цифровыми средствами,

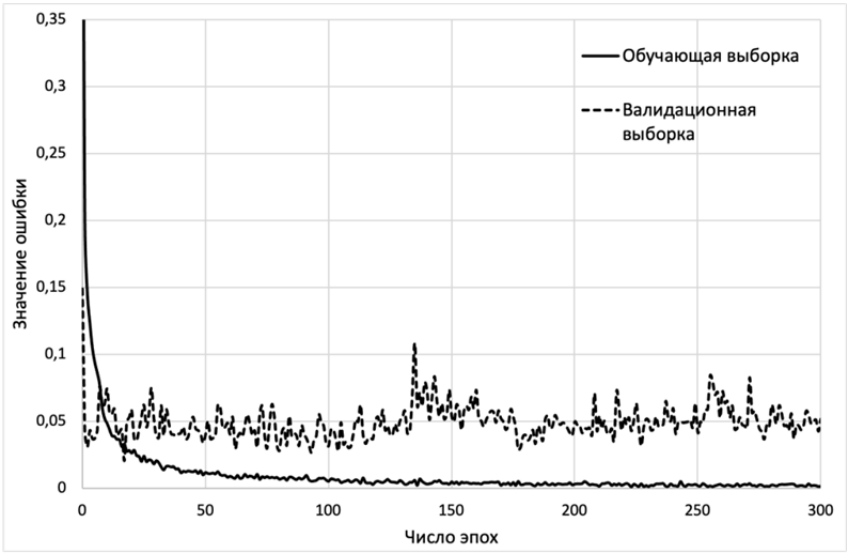

Рис. 3. Изменение средней ошибки на обучающей и валидационной выборках в зависимости от числа эпох

использовался в том числе и набор, содержащий подмену выражений лиц. Примеры изображения из каждой категории набора данных FaceForensics++ представлены на Рис. 4.

По результатам экспериментов построены ROC-кривые для каждого типа манипуляции (Рис. 5), для построения которых оценивались следующие метрики:

1. Вероятность истинно-положительной классификации:

$$TPR = \frac{\text{число оригинальных лиц, для которых классификатор решил,что это оригиналы}}{\text{общее число поддельных лиц}} \quad (3)$$

2. Вероятность ложно-положительной классификации:

$$FPR = \frac{\text{число поддельных лиц, для которых классификатор решил, что это оригиналы}}{\text{общее число поддельных лиц}} \quad (4)$$

В Табл. 2 приведены результаты сравнения предложенного метода с существующими подходами для каждого типа манипуляций. В качестве критерия сравнения использовалась метрика точность (Accuracy), вычисляемая по формуле:

$$Acc = \frac{\text{число верных классификаций}}{\text{общее число лиц}} \quad (5)$$

Анализируя полученные результаты, можно сделать вывод о том, что предложенные архитектура, гиперпараметры и препроцессинг, позволяют с высокой точностью обнаруживать манипуляции с изображениями лиц цифровыми средствами, что в свою очередь может повысить защищенность системы лицевых биометрических систем.

4. Применение предложенного метода для повышения защищенности лицевых биометрических систем

Согласно ГОСТ Р 54412-2019 (ISO/IEC TR 24741:2018) биометрическая система — это система, предназначенная для автоматического распознавания индивидов, основанном на их

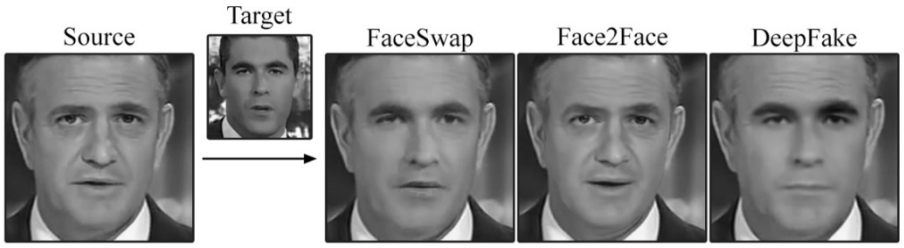

Рис. 4. Пример изображений из категорий FaceSwap, Face2Face и DeepFake набора FaceForensics++

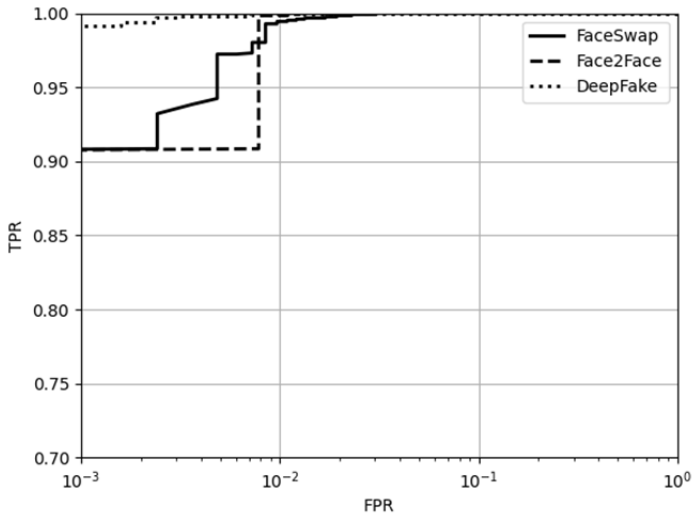

Рис. 5. ROC-кривые для предложенного метода обнаружения манипуляций для каждого типа манипуляций из базы FaceForensics++

Табл. 2. Сравнение точности обнаружения манипуляций с лицами для различных типов манипуляций

Метод	FaceSwap	DeepFake	Face2Face
MesoNet [11]	98.15	99.24	98.35
Bayar [31]	96.8	99.25	99.04
XceptionNet [10]	98.39	99.29	99.23
Khodabakhsh [32]	99.11	99.3	98.25
Jin [16]	-	98.01	93.94
Lu [17]	**99.53**	99.31	**99.69**
Представленный метод	99.28	**99.76**	99.48

биологических и поведенческих характеристиках. Таким образом, лицевая биометрическая система предназначена для автоматического распознавания индивидов на основе биологических и поведенческих характеристик, полученных по изображению лица.

Согласно тому же нормативному документу, биометрические устройства и системы могут быть подвергнуты различным испытаниям, которые могут включать в себя оценку: эксплуатационных характеристик (в терминах вероятностей ошибок и производительности); надежности, доступности и удобства эксплуатации; степени защищенности; безопасности; приемлемости системы для пользователя; влияния человеческих факторов; коэффициента эффективности затрат; степени соответствия правилам конфиденциальности.

Предложенный в работе метод предназначен для повышения степени защищенности лицевых биометрических систем. На Рис. 6 показана структура биометрической системы общего вида (ГОСТ Р 54412-2019 (ISO/IEC TR 24741:2018)). На каждый из компонентов системы могут быть осуществлены различные типы атак: на канал связи, на базу данных, на захват данных, их обработку и принятие решения. Все они, за исключением атак на устройство ввода, являются общими для различных модальностей (лицо, голос, отпечатки пальца). Для противодействия этим атакам применяются методы цифрового кодирования и шифрования канала связи. Компонент системы, отвечающий за получение биометрического образца, подвержен в том числе атакам спуфинга — попытке подмены биометрической характеристики. При этом злоумышленники могут использовать как распечатанную фотографию, так и модификацию изображения лица цифровыми средствами. Именно на обнаружение попыток использования злоумышленником образца, полученного цифровыми средствами манипуляции с изображениями лиц и направлен представленный метод.

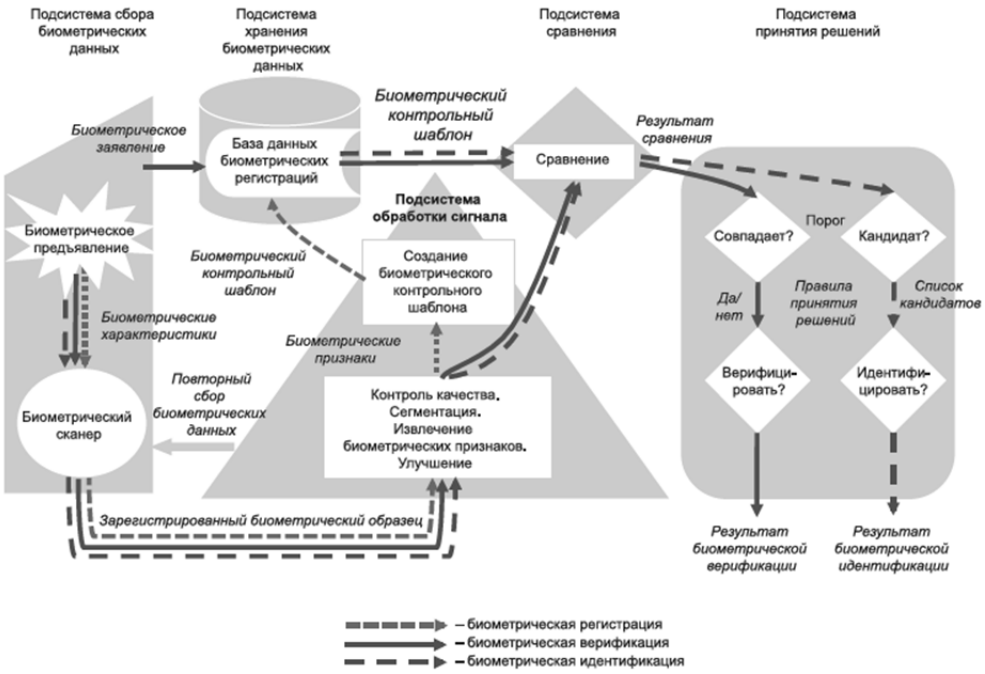

Рис. 6. Компоненты биометрической системы общего вида

Для повышения устойчивости лицевой биометрической системы к атакам спуфинга на модуль ввода биометрической информации на этапе контроля качества проводится проверка полученного биометрического образца на то, является ли этот образец поддельным и не попытался ли злоумышленник его подделать. Таким образом, представленный в работе метод классификации, является ли изображение лица поддельным или нет, используется на этапе анализа биометрического образца до проведения этапа извлечения биометрических признаков.

В случае, если полученный биометрический образец признается поддельным, биометрическая регистрация, верификация и идентификация прекращаются. А, следовательно, поддельный образец не попадает ни в подсистему хранения данных, ни в подсистему сравнения и принятия решений. Следовательно, использование модуля определения применения технологии манипуляций с изображениями лиц на этапе контроля качества биометрического образца, позволяет находить и пресекать попытки маскировки под авторизованного клиента, используя цифровые методы манипуляции с идентичностью лиц, что повышает степень защищенности лицевой биометрической системы.

Заключение

В работе исследована задача повышения устойчивости систем биометрической аутентификации по изображению лица к атакам на модуль ввода биометрической информации цифровыми средствами манипуляции с идентичностью лиц.

Предложенное решение основывается на классификации признаков, полученных сверточной нейронной сетью для предварительно обработанного изображения лица. Проведенный вычислительный эксперимент показал, что предложенный метод достигает точности не менее 99% для различных способов манипуляции с лицами. Результаты экспериментов также показывают эффективность предложенного подхода при сравнении его с другими известными методами, протестированными на том же наборе данных.

Дальнейшее развитие метода может быть направлено на комбинацию различных классификаторов, а также использованию нескольких кадров видеопоследовательности.

Литература

1. Генеральная П.Р.Ф. Главное управление правовой статистики и информационных технологий. Состояние преступности в России за январь-октябрь 2021 г.

//Официальный сайт Портала правовой статистики Генеральной Прокуратуры РФ. URL: http://crimestat.ru/analytics. 2021.
2. Thies J. et al. Face2face: Real-time face capture and reenactment of rgb videos //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. С. 2387-2395.
3. Tolosana R. et al. Deepfakes and beyond: A survey of face manipulation and fake detection //Information Fusion. 2020. Т. 64. С. 131-148.
4. Korshunova I. et al. Fast face-swap using convolutional neural networks //Proceedings of the IEEE international conference on computer vision. 2017. С. 3677-3685.
5. Agarwal A. et al. Swapped! digital face presentation attack detection via weighted local magnitude pattern //2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE. 2017. С. 659-665.
6. Mo H., Chen B., Luo W. Fake faces identification via convolutional neural network //Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. 2018. С. 43-47.
7. Agarwal S. et al. Protecting World Leaders Against Deep Fakes //CVPR workshops. – 2019. Т. 1.
8. Li X. et al. Fighting against deepfake: Patch&pair convolutional neural networks (PPCNN) //Companion Proceedings of the Web Conference 2020. – 2020. – С. 88-89.
9. Li L. et al. Face x-ray for more general face forgery detection //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. С. 5001-5010.
10. Rossler A. et al. Faceforensics++: Learning to detect manipulated facial images //Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019. С. 1-11.
11. Afchar D. et al. Mesonet: a compact facial video forgery detection network //2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE. 2018. С. 1-7.
12. Ding X. et al. Swapped face detection using deep learning and subjective assessment //EURASIP Journal on Information Security. 2020. Т. 2020. С. 1-12.
13. Do N. T., Na I. S., Kim S. H. Forensics face detection from GANs using convolutional neural network //2018 International Symposium on Information Technology Convergence (ISITC 2018). South Korea. 2018.
14. Hsu C. C., Zhuang Y. X., Lee C. Y. Deep fake image detection based on pairwise learning //Applied Sciences. 2020. Т. 10. №. 1. С. 370.
15. Fernando T. et al. Exploiting human social cognition for the detection of fake and fraudulent faces via memory networks //arXiv preprint arXiv:1911.07844. 2019.
16. Jin X., Ye D., Chen C. Countering Spoof: Towards Detecting Deepfake with Multidimensional Biological Signals //Security and Communication Networks. 2021. Т 2021.
17. Lu Y. et al. Channel-Wise Spatiotemporal Aggregation Technology for Face Video Forensics //Security and Communication Networks. 2021. Т. 2021.
18. Wang R. et al. Fakespotter: A simple baseline for spotting ai-synthesized fake faces. arXiv 2019 //arXiv preprint arXiv:1909.06122.
19. Fernandes S. et al. Detecting deepfake videos using attribution-based confidence metric //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020. С. 308-309.
20. Khodabakhsh A., Busch C. A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling //2020 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE. 2020. С. 1-5.
21. Volkova S. S., Bogdanov A. S. A deep learning approach to face swap detection //International Journal of Open Information Technologies. 2021. Т. 9. №. 10. С. 16-20.
22. Tarasiou M., Zafeiriou S. Extracting deep local features to detect manipulated images of human faces //2020 IEEE International Conference on Image Processing (ICIP). – IEEE, 2020. С. 1821-1825.
23. Tang X. et al. Pyramidbox: A context-assisted single shot face detector //Proceedings of the European conference on computer vision (ECCV). 2018. С. 797-813.
24. Xiong X., De la Torre F. Supervised descent method and its applications to face alignment //Proceedings of the IEEE conference on computer vision and pattern recognition. 2013. С. 532-539.
25. Sandler M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. С. 4510-4520.
26. Deng J. et al. Imagenet: A large-scale hierarchical image database //2009 IEEE conference on computer vision and pattern recognition. Ieee. 2009. С. 248-255.
27. Defazio A., Jelassi S. Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization //arXiv preprint arXiv:2101.11075. 2021.
28. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization //Journal of machine learning research. 2011. Т. 12. №. 7.
29. Kingma D. P., Ba J. Adam: A method for stochastic optimization //arXiv preprint arXiv:1412.6980. 2014.
30. Loshchilov I., Hutter F. SGDR: Stochastic gradient descent with warm restarts //International Conference on Learning Representations. 2017.
31. Bayar B., Stamm M. C. A deep learning approach to universal image manipulation detection using a new convolutional layer //Proceedings of the 4th ACM workshop on information hiding and multimedia security. 2016. С. 5-10.
32. Khodabakhsh A., Busch C. A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling //2020 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE. 2020. С. 1-5.

Волкова Светлана Сергеевна. Кандидат технических наук. Доцент кафедры прикладной математики. ФГБОУ ВО «Вологодский государственный университет». Области исследований: машинное обучение, компьютерное зрение. E-mail: malysheva.svetlana.s@gmail.com

Method for DeepFake Detection Using Convolutional Neural Networks

S. S. Volkova

Vologda State University, Vologda, Russia

Abstract. The article proposed the face anti-digital-spoofing countermeasures method for improving the protection of the facial biometric system. The DeepFake detection method is based on the convolutional neural networks, trained on a large dataset that contains different fake types with different qualities. This has resulted in at least 99% of detection quality. The suggested method can be used to increase the protection of facial biometric systems by reducing the risk of unauthorized access.
Keywords: biometric authentication, spoofing, DeepFake, convolutional neural network, recognition.

DOI 10.14357/20718594220206

References

1. The Prosecutor General's Office of the Russian Federation. Department of legal statistics, information technology. 2021. Sostoyanie prestupnosti v Rossii za yanvar-octyabr 2021 [The crime rate in Russia for january-october 2021] // Official portal of legal statistics department URL: http://crimestat.ru/analytics.
2. Thies J. et al. Face2face: Real-time face capture and reenactment of rgb videos //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016. P. 2387-2395.
3. Tolosana R. et al. Deepfakes and beyond: A survey of face manipulation and fake detection //Information Fusion. 2020. T. 64. P. 131-148.
4. Korshunova I. et al. Fast face-swap using convolutional neural networks //Proceedings of the IEEE international conference on computer vision. 2017. P. 3677-3685.
5. Agarwal A. et al. Swapped! digital face presentation attack detection via weighted local magnitude pattern //2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE/ 2017. P. 659-665.
6. Mo H., Chen B., Luo W. Fake faces identification via convolutional neural network //Proceedings of the 6th ACM Workshop on Information Hiding and Multimedia Security. 2018. P. 43-47.
7. Agarwal S. et al. Protecting World Leaders Against Deep Fakes //CVPR workshops. – 2019. T. 1.
8. Li X. et al. Fighting against deepfake: Patch&pair convolutional neural networks (PPCNN) //Companion Proceedings of the Web Conference 2020. – 2020. – P. 88-89.
9. Li L. et al. Face x-ray for more general face forgery detection //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020. P. 5001-5010.
10. Rossler A. et al. Faceforensics++: Learning to detect manipulated facial images //Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2019. P. 1-11.
11. Afchar D. et al. Mesonet: a compact facial video forgery detection network //2018 IEEE International Workshop on Information Forensics and Security (WIFS). IEEE. 2018. P. 1-7.
12. Ding X. et al. Swapped face detection using deep learning and subjective assessment //EURASIP Journal on Information Security. 2020. T. 2020. P. 1-12.
13. Do N. T., Na I. S., Kim S. H. Forensics face detection from GANs using convolutional neural network //2018 International Symposium on Information Technology Convergence (ISITC 2018). South Korea. 2018.
14. Hsu C. C., Zhuang Y. X., Lee C. Y. Deep fake image detection based on pairwise learning //Applied Sciences. 2020. T. 10. №. 1. P. 370.
15. Fernando T. et al. Exploiting human social cognition for the detection of fake and fraudulent faces via memory networks //arXiv preprint arXiv: 1911.07844. 2019.
16. Jin X., Ye D., Chen C. Countering Spoof: Towards Detecting Deepfake with Multidimensional Biological Signals //Security and Communication Networks. 2021. T. 2021.
17. Lu Y. et al. Channel-Wise Spatiotemporal Aggregation Technology for Face Video Forensics //Security and Communication Networks. 2021. T. 2021.
18. Wang R. et al. Fakespotter: A simple baseline for spotting ai-synthesized fake faces. arXiv 2019 //arXiv preprint arXiv:1909.06122.
19. Fernandes S. et al. Detecting deepfake videos using attribution-based confidence metric //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. 2020. P. 308-309.
20. Khodabakhsh A., Busch C. A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling //2020 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE. 2020. P. 1-5.
21. Volkova S. S., Bogdanov A. S. A deep learning approach to face swap detection //International Journal of Open Information Technologies. 2021. T. 9. №. 10. P. 16-20.
22. Tarasiou M., Zafeiriou S. Extracting deep local features to detect manipulated images of human faces //2020 IEEE International Conference on Image Processing (ICIP). – IEEE. 2020. P. 1821-1825.
23. Tang X. et al. Pyramidbox: A context-assisted single shot face detector //Proceedings of the European conference on computer vision (ECCV). 2018. P. 797-813.
24. Xiong X., De la Torre F. Supervised descent method and its applications to face alignment //Proceedings of the IEEE conference on computer vision and pattern recognition. 2013. P. 532-539.

25. Sandler M. et al. Mobilenetv2: Inverted residuals and linear bottlenecks //Proceedings of the IEEE conference on computer vision and pattern recognition. 2018. P. 4510-4520.
26. Deng J. et al. Imagenet: A large-scale hierarchical image database //2009 IEEE conference on computer vision and pattern recognition. Ieee. 2009. P. 248-255.
27. Defazio A., Jelassi S. Adaptivity without compromise: a momentumized, adaptive, dual averaged gradient method for stochastic optimization //arXiv preprint arXiv:2101.11075. 2021.
28. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization //Journal of machine learning research. 2011. Т. 12. №. 7.
29. Kingma D. P., Ba J. Adam: A method for stochastic optimization //arXiv preprint arXiv:1412.6980. 2014.
30. Loshchilov I., Hutter F. SGDR: Stochastic gradient descent with warm restarts //International Conference on Learning Representations. 2017.
31. Bayar B., Stamm M. C. A deep learning approach to universal image manipulation detection using a new convolutional layer //Proceedings of the 4th ACM workshop on information hiding and multimedia security. 2016. P. 5-10.
32. Khodabakhsh A., Busch C. A Generalizable Deepfake Detector based on Neural Conditional Distribution Modelling //2020 International Conference of the Biometrics Special Interest Group (BIOSIG). IEEE. 2020. P. 1-5.

Volkova Svetlana S. Candidate of technical sciences. Associate professor, Department of applied mathematics, Vologda State University. Research areas: machine learning, computer vision. E-mail: malysheva.svetlana.s@gmail.com