

Применение легковесной сиамской нейросети для формирования вектора признаков в системе васкулярной аутентификации

Д.Е. Прозоров¹, А.В. Земцов¹

¹ Вятский государственный университет, 610000, Россия, г. Киров, ул. Московская, д. 36

Аннотация

В статье анализируется возможность использования сиамской сверточной нейросети для решения задачи васкулярной аутентификации на встраиваемой аппаратной платформе с ограниченными вычислительными ресурсами (Orange Pi One). Выполнен краткий обзор современных методов вычисления векторов признаков изображений, применяемых в задачах классификации, сравнения или поиска изображений по контенту: на основе вариационных рядов (гистограмм), локальных дескрипторов, дескрипторов особых точек, дескрипторов на основе хэш-функций, нейросетевых дескрипторов. Предложена архитектура биометрической системы аутентификации по изображениям ладоней в видимом и ближнем ИК-спектрах на основе сиамской сверточной нейросети. Разработанное программное решение позволяет использовать сиамскую нейросеть в режимах «полная сеть» (используются оба симметричных канала нейросети) и «половина нейросети» (используется только один канал) для сокращения времени сравнения векторов биометрических данных зарегистрированных пользователей биометрической системы аутентификации. Показаны преимущества нейросетевых признаков, заключающиеся в универсальности, масштабируемости и конкурентоспособности, в том числе на встраиваемых аппаратно-программных решениях с ограниченными вычислительными ресурсами при отсутствии графических ускорителей. Исследования показали возможность повышения качества классификации изображений ладоней с 0,929 до 0,968 по метрике «overall accuracy» при использовании сиамской нейросети вместо метода перцептивного хэширования при сопоставимом времени определения зарегистрированной в биометрической системе аутентификации персоны. В экспериментах осуществлялся поиск по базе данных из 2000 изображений для 400 персон.

Ключевые слова: биометрическая аутентификация, обработка изображений, дескрипторы изображений, искусственная нейронная сеть, сиамская нейронная сеть.

Цитирование: Прозоров, Д.Е. Применение легковесной сиамской нейросети для формирования вектора признаков в системе васкулярной аутентификации / Д.Е. Прозоров, А.В. Земцов // Компьютерная оптика. – 2023. – Т. 47, № 3. – С. 433–441. – DOI: 10.18287/2412-6179-CO-1204.

Citation: Prozorov DE, Zemtsov AV. Using a lightweight Siamese neural network for generating a feature vector in a vascular authentication system. Computer Optics 2023; 47(3): 433-441. DOI: 10.18287/2412-6179-CO-1204.

Введение

Системы биометрической аутентификации (БАС) получили широкое распространение, и их количество постоянно увеличивается. В качестве факторов аутентификации в БАС используются индивидуальные биометрические данные, однозначно характеризующие пользователя системы: изображение лица, отпечаток пальца, фрагмент речевого сигнала и др. В системах васкулярной аутентификации анализируется рисунок вен, формируемый в результате обработки изображения ладони руки. Вследствие использования сканеров инфракрасного диапазона для получения первичных изображений, методы васкулярной аутентификации более защищены от подделки биометрических данных и обеспечивают меньший уровень ошибок в процессе принятия решений по сравнению, например, с методами аутентификации на основе папиллярного узора пальцев.

Первичные изображения, получаемые со сканера пропускного пункта БАС, неудобны для прямого ис-

пользования в алгоритмах классификации, применяемых в работе БАС: они занимают большой объем памяти и часто содержат избыточную долю информации, не оказывающей существенного влияния на качество решения. Поэтому в процессе предварительной обработки первичные изображения преобразуются в более компактный вид – матрицу или вектор вещественных чисел (вектор признаков), позволяющий однозначно идентифицировать пользователя БАС.

Зачастую вычисление «хорошего» вектора признаков для описания биометрических данных требует применения вычислительно-затратных методов обработки и входит в противоречие с возможностями аппаратной платформы. Особенно остро эта проблема встает при реализации БАС на дешевых встраиваемых аппаратных платформах.

Таким образом, актуальной задачей при разработке мобильных и встраиваемых систем васкулярной аутентификации является задача вычисления вектора признаков биометрических данных в условиях ограниченных вычислительных ресурсов. С этой целью в

данной работе предлагается использовать метод на основе сиамской нейронной сети, позволяющий найти компромисс между производительностью и показателями качества аутентификации БАС.

1. Методы вычисления вектора признаков

Научное сообщество разработало довольно много способов вычисления вектора признаков, используемых в задачах обработки изображений. Существенная часть из них может быть использована и в задачах биометрической аутентификации.

1.1. Вариационный ряд (гистограмма)

Идея использования гистограммы ориентированных градиентов была высказана в патенте R.K. McConnell [1] и позже использована Mitsubishi Electric Research Laboratories в проекте по распознаванию жестов рук [2]. Существуют разновидности алгоритма HOG для описания прямоугольных (R-HOG) и круглых (C-HOG) фрагментов изображений. Вариантом HOG является вектор параметров «Pyramid Histogram of Oriented Gradients» (PHOG), позволяющий описывать фрагменты изображения с разным уровнем масштаба [3].

1.2. Локальные дескрипторы

T. Ojala, M. Pietikhenl и D. Harwood предложили для сравнения изображений использовать т.н. «локальный дескриптор» [4]. Базовый метод заключается в конструировании вектора признаков – локального бинарного паттерна (LBP) – путем расчета некоторой метрики для пикселей локальной области изображения, последовательно обрабатываемых по часовой или против часовой стрелки. Часто указанная метрика является функцией интенсивности (яркости) пикселей.

Существует довольно много модификаций базового метода LBP.

Комбинация метода CS-LDP (center-symmetric local derivative pattern) с CS-LBP (center-symmetric local binary pattern) образует метод вычисления признаков HCS-LP (Hybrid center-symmetric local pattern) [5]. Оператор LBP инвариантен к изменениям освещенности и может применяться во многих приложениях обработки изображений. CS-LBP, в отличие от LBP, оперирует почти в два раза меньшим количеством сравнений за счет использования симметричных пар точек. Вектор признаков CS-LDP имеет ту же длину, что и CS-LBP, но собирает более детальную информацию о локальной области за счет выполнения сравнений как между симметричными парами точек на окраинах локальной области, так и между парами точек центр – окраина. Вектор HCS-LP просто объединяет признаки CS-LBP и CS-LDP.

Базовый метод LBP слабо устойчив к воздействию шумов на вектор локальных признаков. В работе [6] предложен CS-LTP (Center Symmetric Local Ternary

Pattern), частично устраняющий данный недостаток. CS-LTP является обобщением CS-LBP и может адаптироваться к локальным изменениям яркости на изображении. Показано, что сочетание CS-LTP признаков с гистограммой относительной интенсивности (HRI) дает более высокие результаты. Также менее чувствительный к шуму дескриптор XCS-LBP (eXtended Center-Symmetric Local Binary Pattern), являющийся расширенным вариантом CS-LBP, предложен в работе [7]. XCS-LBP формирует такой же короткий вектор признаков, как CS-LBP.

Как CS-LBP, так и CS-LTP признаки формируются без учета интенсивности центрального пикселя локальной области. В eCS-LTP (Extended Center-Symmetric Local Ternary Pattern) – расширенном варианте оператора CS-LTP – был устранен этот недостаток при сохранении таких особенностей LTP, как простота, короткий вектор признаков и устойчивость к вариациям освещенности и поворотам [8].

В отличие от CS-LBP метод CS-LMP [9] не использует бинарное квантование различий между пикселями, что позволяет учитывать нюансы изменения яркости в локальной области. Формируемый вектор признаков инвариантен к масштабированию изображения, повороту, освещенности объекта и изменению точки зрения.

Обобщение CS-LTP и CS-LMP позволило разработать метод CS-LTMP (Center Symmetric Local Ternary Mapped Patterns) [10]. Комбинируя полученный вектор признаков CS-LTP с XCS-LMP оператором (eXtended Center Symmetric Local Mapped Pattern), авторы получили XCS-LTMP (eXtended Center Symmetric Local Ternary Mapped Pattern) вектор признаков. Показано, что CS-LTMP и XCS-LTMP признаки превосходят CS-LBP, CS-LDP, CS-LMP, CS-LDMP, XCS-LBP, XCS-LMP, CS-LTP и eCS-LTP при решении задачи многоклассовой классификации по метрике accuracy на датасете CIFAR10.

Описанные выше признаки подходят для обработки изображений в градациях серого, наиболее часто используемых при решении задач аутентификации по рукам (отпечаткам пальцев, папиллярному узору или по венам). Существуют также признаки, пригодные для параметризации цветных изображений, что может быть полезно в тех случаях, когда цветовая информация существенно повышает качество решения задачи. К таким признакам, например, относятся CEDD (Color and Edge Directivity Descriptor) [11], FCTH (Fuzzy Color and Texture Histogram Descriptor) [12], CLD (Color Layout Descriptor) и объединенный дескриптор JCD (Joint Composite Descriptor) [13].

1.3. Дескрипторы особых точек

Помимо вышеперечисленных, в системах поиска изображений по запросу (Context-Based Search, CBS) широко применяются методы поиска и дескрипторы «особых точек» (keypoints).

Известным методом формирования векторов признаков на основе «ключевых» точек является SIFT (Scale-Invariant Feature Transform) [14]. Метод запатентован университетом Британской Колумбии. SIFT-дескрипторы обладают устойчивостью к ограниченным аффинным преобразованиям, изменениям масштаба и поворотам. SIFT является одним из наиболее эффективных, но при этом вычислительно затратных методов.

Во многих случаях при сравнении изображений по «ключевым» точкам имеется возможность достижения компромисса между вычислительной сложностью метода, объемом памяти используемой базой данных дескрипторов, вычислительной сложностью метода и скоростью сравнения изображений. Так, для сокращения объема базы, содержащей дескрипторы изображений (что важно при построении мобильных или встраиваемых систем), можно использовать дескрипторы SURF, PCA-SIFT, A-KAZE, ORB или BRISK.

SURF (Speeded Up Robust Features) инвариантен к масштабированию и вращению, менее эффективен в сравнении с SIFT, но работает значительно быстрее [15]. В PCA-SIFT для сокращения размера вектора признаков используется т.н. «метод главных компонент» (PCA) [16]. Вариантом SIFT-дескриптора является GLOH (Gradient Location and Orientation Histogram) [17], разработанный для повышения устойчивости и различимости дескрипторов. BRIEF (Binary Robust Independent Elementary Features) [18] и ORB (Oriented FAST and rotated BRIEF) [19] позиционируются как быстрая и эффективная замена SIFT-дескрипторам. ORB формирует очень короткий вектор признаков, показывает сравнимые с SIFT результаты на зашумленных изображениях и может конкурировать с SIFT при углах поворота, кратных 90 градусам. A-KAZE (Accelerated-KAZE) [20] инвариантен к изменению масштаба изображения, поворотам и частично к аффинным преобразованиям. Формирует вектор признаков, сравнимый с BRISK и меньший, чем SIFT. При этом вычислительно эффективнее SIFT и SURF. BRISK (Binary Robust Invariant Scalable Keypoints) – очень быстрый алгоритм с небольшим вектором признаков, превышающий производительность SURF в несколько раз [21]. Устойчив к изменениям масштаба и поворотам; уступает по эффективности SIFT, но превосходит ранее перечисленные дескрипторы во многих случаях (аффинные преобразования, сильные повороты, сильные изменения масштаба). В целом, наиболее устойчивыми к изменениям масштаба являются дескрипторы SIFT, SURF и BRISK; ORB, BRISK и A-KAZE показывают хорошие результаты при поворотах; дескрипторы ORB и BRISK обеспечивают наиболее быстрое сравнение изображений [22].

1.4. Дескрипторы на основе хэш-функций

При решении задач параметризации изображений в приложениях с ограниченной памятью и вычисли-

тельными ресурсами могут оказаться полезными методы на основе перцепционного хеширования. В то время как криптографическое хеширование должно давать радикально различный результат при малом изменении входных данных, функции преобразования изображений должны давать близкие (в линейном или нелинейном пространстве) вектора для похожих объектов на изображениях.

Методы хеширования, применимые в задачах поиска похожих изображений, известны довольно давно, являются очень быстрыми, но в настоящий момент уступают по эффективности более современным, на основе ключевых точек или нейросетевых дескрипторов. Практически все перцептивные хэш-дескрипторы устойчивы к масштабированию изображений.

Одним из самых простых способов отображения изображения в вектор признаков является метод порогового усреднения соседних пикселей (цветовая шкала изображения предварительно переводится в градации серого). На этой идее основан алгоритм Simple Hash (Average Hash). Average Hash очень быстр, устойчив к изменению яркости, контраста и манипуляциям с цветами изображения. Дифференциальный хэш (dHash) по аналогии с Average Hash преобразует изображение в градации серого и использует сравнение групп пикселей слева-направо (или сверху-вниз).

Возможно, наиболее известным является метод [23] и open-source библиотека [24] перцепционного хеширования pHash. Данный алгоритм несколько более сложен по сравнению с Average Hash: использует масштабирование исходного изображения, дискретное косинусное преобразование для отображения данных в частотную область, обрезку результата для учета только средне- и низкочастотной части (эти частоты характеризуют структуру изображения) и медианного (или обычного) усреднения. Хэш на основе вейвлет-преобразования wHash отличается использованием двумерных вейвлет-функций вместо дискретного косинусного преобразования.

В том случае, если важно сохранение цветовой информации, можно использовать Color Hash. Color Hash формирует вектор признаков путем кодирования данных яркости, цветности и насыщенности пикселей.

Быстрый способ сравнения перечисленных хэш-дескрипторов заключается в использовании расстояния Хэмминга. Реализации указанных хэш-функций доступны, например, в python-ориентированной библиотеке ImageHash [25].

Следует отметить, что векторы признаков, пригодные для решения задачи биометрической аутентификации, могут быть получены путем использования любой комбинации вышеописанных дескрипторов.

1.5. Нейросетевые дескрипторы

Как видно из предыдущих параграфов, имеется большое разнообразие методов вычисления векторов

признаков изображений. Традиционно исследователи при построении систем поиска схожих изображений (либо для решения других, аналогичных задач) проходят этап выбора признаков, специфичных для конкретных рассматриваемых классов изображений. Результатом может стать комбинация нескольких дескрипторов и «сжатие» полученного вектора с целью декорреляции с помощью, например, метода главных компонент (principal component analysis, PCA). Широко известным методом построения поискового индекса на основе отобранных признаков является «мешок визуальных слов» [26]. Метод использует кластеризацию признаков для формирования описания «типичного» представителя класса.

Зачастую целью этого этапа является формирование вектора признаков, позволяющего максимально дистанцировать объекты разных классов в линейном или нелинейном пространстве признаков при сохранении компактности самого вектора признаков. Такой подход позволяет уменьшить количество ошибок различения объектов, ускорить их обработку и сократить требования к вычислительным ресурсам системы обработки изображений. При этом выбор набора признаков может зависеть от решаемой задачи. Так, при решении задач обработки изображений в задачах классификации подписей сотрудников на документах, изображений лиц или ладоней со сканеров систем безопасности принципы вычисления наилучших векторов признаков могут сильно отличаться.

С точки зрения универсализации процесса поиска наилучшего вектора признаков вызывают интерес нейросетевые методы и дескрипторы, формирующие-ся в процессе обучения нейросетевых моделей.

Подавляющее большинство нейросетевых моделей, применяемых для обнаружения, классификации, трекинга и прочих задач обработки изображений, содержат в своей структуре сверточные слои, которые по сути являются фильтрами характерных визуальных особенностей объектов на изображениях. Архитектура типичной сверточной нейросети представлена на рис. 1 и содержит упрощенно часть 1, предназначенную для формирования карт признаков, и часть 2, предназначенную для кодирования признаков и решения задачи классификации. Часть 1, помимо сверточных слоев, может содержать слои нормализации (batch normalization), субдискретизации (pooling) и регуляризации (dropout). Часть 2 в простейшем случае содержит полностью связанные слои (dense), «сжимающие» карты признаков в итоговый вектор, на основании нормализованных значений которого принимается решение о классе объекта.

В настоящее время известно большое количество архитектур сверточных нейросетей (DenseNet, EfficientNet, InceptionResNetV2, InceptionV3, MobileNet, MobileNetV2, MobileNetV3, NASNet, ResNet, ResNetV2, VGG, Xception и др.), позволяющих достичь высокого качества классификации объ-

ектов интереса. Одним из немногих недостатков перечисленных нейросетей является сложность их практического применения во встраиваемых и мобильных приложениях на базе, например, низкопроизводительных ARM-архитектур без графических ускорителей. Кроме того, классические нейросетевые классификаторы со структурой, представленной на рис. 1, неудобны для решения задач биометрической аутентификации, поскольку требуют переобучения после регистрации новых пользователей БАС (регистрация нового пользователя влечет за собой изменение количества классов, которые должен различать классификатор).

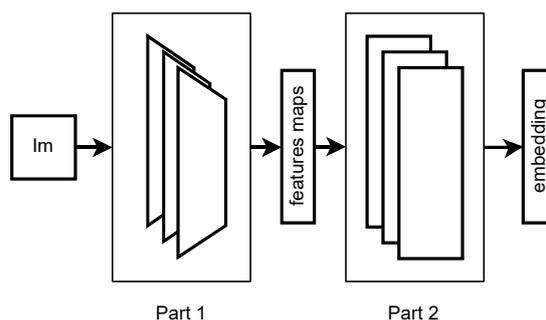


Рис. 1. Упрощенная структура сверточной нейросети

Одним из решений, позволяющим использовать преимущества нейросетевых решений в случае нефиксированного количества классов объектов интереса, являются т.н. «сиамские» нейросети.

2. Сиамская нейросеть

Сиамская нейросеть – это архитектура, в которой используются идентичные веса нейросетей в двух симметричных каналах обработки входных данных (рис. 1) [27]. Целью такой архитектуры является получение одинаковых векторов признаков для одинаковых наборов данных на входах каналов нейросети. Формируемые вектора признаков используются для вычисления меры сходства между исходными объектами. По результатам этих вычислений входные изображения классифицируются как относящиеся к одному либо к разным классам. Метод относится к классу One-Shot Learning (обучение с первого раза) и не требует в процессе обучения большого количества примеров для каждого класса. Недостатком является почти двухкратное увеличение количества вычислений по сравнению с аналогичной сверточной сетью (рис. 1).

Упрощенная архитектура сиамской нейросети представлена на рис. 2.

В качестве функции потерь при обучении сиамских сетей часто используется Contrastive Loss (1) [28].

$$L\left(W, \left(Y, \overline{X_1}, \overline{X_2}\right)^i\right) = (1 - Y) \frac{1}{2} (D_w)^2 + (Y) \frac{1}{2} \left\{ \max(0, m - D_w) \right\}^2, \tag{1}$$

$$D_W(\bar{X}_1, \bar{X}_2) = \|G_W(\bar{X}_1) - G_W(\bar{X}_2)\|_2. \quad (2)$$

В формуле (1): $(Y, \bar{X}_1, \bar{X}_2)^i$ – i -й триплет из пары входных векторов \bar{X}_1, \bar{X}_2 и метки Y , где $Y=0$, означает, что объекты, закодированные векторами \bar{X}_1 и \bar{X}_2 , принадлежат одному классу, а $Y=1$ – разным классам; $D_W(\bar{X}_1, \bar{X}_2)$ – дистанция между выходами $G_W(\bar{X}_1)$ и $G_W(\bar{X}_2)$ (в данном случае – евклидова); $m > 0$ – отступ (правая часть выражения (1) – параметр «несхожести» – рассчитывается только для объектов на дистанции менее радиуса m).

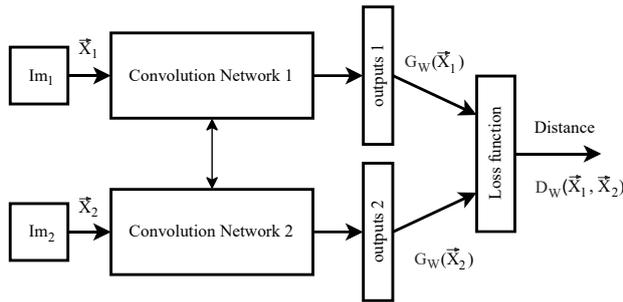


Рис. 2. Упрощенная архитектура сиамской нейросети в процессе обучения

Упрощенно схема использования сиамской нейросети выглядит следующим образом.

- Нейросетевая модель обучается на примерах, состоящих из пар изображений, принадлежащих одной и той же персоне (рис. 3а, б) либо разным персонам (рис. 3в, г). Количество таких примеров для каждой персоны может быть относительно невелико (десятки или сотни экземпляров).
- Как в процессе обучения, так и в дальнейшем применении вычисляются вектора признаков входных изображений $G_W(\bar{X}_1)$ и $G_W(\bar{X}_2)$.
- На этапе обучения веса нейросети итеративно корректируются с учетом функции потерь $L(W, (Y, \bar{X}_1, \bar{X}_2)^i)$.
- На этапе применения вычисляется только метрика близости (расстояние) между векторами признаков. В случае (1) этой метрикой является евклидово расстояние – норма L2 векторов $G_W(\bar{X}_1)$ и $G_W(\bar{X}_2)$.
- Если вычисленное расстояние меньше заданного порога, то принимается решение о принадлежности изображений одной персоне. В противном случае – разным.

В процессе принятия решений могут возникать ошибки ложного доступа (FAR – False Acceptance Rate – система предоставляет доступ персоне без права доступа) и ошибки ложного отказа (FRR – False Rejection Rate – доступ запрещён пользователю, у ко-

торого есть право доступа). В данном исследовании анализируются ошибки классификации зарегистрированного пользователя по метрике overall accuracy.

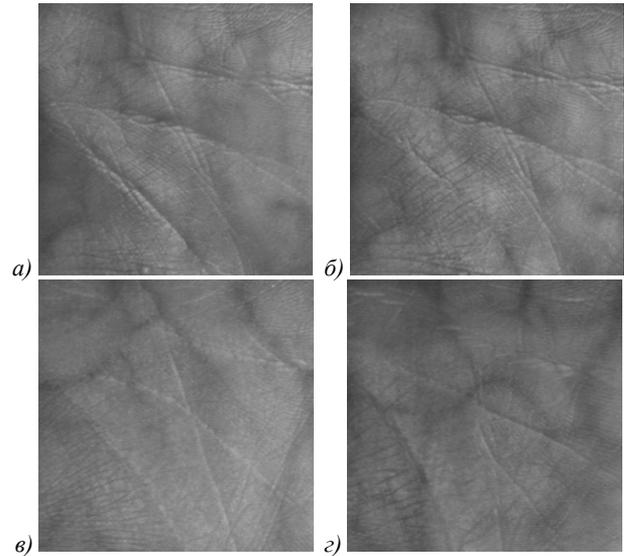


Рис. 3. Фрагменты изображений ладоней в ИК-диапазоне 940 нм, принадлежащие: одной персоне – а), б); разным персонам – в), г). Фрагменты предварительно обработаны и приведены к размеру 105×105 пикселей

Достоинством архитектуры (рис. 2) в системе биометрической аутентификации является возможность оптимизации метрик качества аутентификации при заданных ограничениях на стоимость аппаратного решения и времени аутентификации пользователя. Оптимизация, в частности, производится путем подбора архитектуры сверточной нейросети, используемой в каждом из двух симметричных путей распространения сигнала.

3. Архитектура системы биометрической аутентификации

Как было указано выше, одним из недостатков сиамских нейросетей является почти двухкратное увеличение вычислительной сложности по сравнению с аналогичным классификатором на основе сверточной нейросети. Для сокращения вычислительных затрат на поиск персоны в базе данных предлагается использование сиамской нейросети в двух режимах:

- «полная нейросеть» – соответствует рис. 2 и используется на этапе обучения;
- «половина нейросети» – используется на этапе применения.

В режиме «половина нейросети» один из каналов сиамской нейросети деактивируется. Выход активного канала формирует вектор признаков $G_W(\bar{X}_i)$ экземпляра изображения, поступившего на вход ИНС. Этот вектор признаков последовательно сравнивается со всеми заранее вычисленными шаблонами $H_j = G_W(\bar{X}_j)$, хранящимися в базе данных БАС (множество шаблонов \bar{H} формируется в процессе регистрации пользователей в БАС). Аутентификация

считается успешной, если дистанция между сравниваемыми векторами признаков менее порога принятия решения γ :

$$D_W(\vec{X}_i, \vec{H}_j) = \|G_W(\vec{X}_i) - \vec{H}_j\|_2 < \gamma. \quad (3)$$

Порог γ подбирается исходя из требований к величинам FAR и FRR.

С учетом предложенной схемы работы ИНС процедура биометрической аутентификации включает следующие необходимые шаги.

- Получение изображения со сканера и очистка данных.
- Вычисление и кодирование области интереса (ROI) в виде вектора признаков \vec{X}_i .
- Вычисление эмбединга $G_W(\vec{X}_i)$ для полученной области интереса.
- Определение факта наличия либо отсутствия схожего с $G_W(\vec{X}_j)$ эмбединга в БД БАС.
- Предоставление доступа либо отказ в доступе в зависимости от результатов выполнения предыдущего шага.

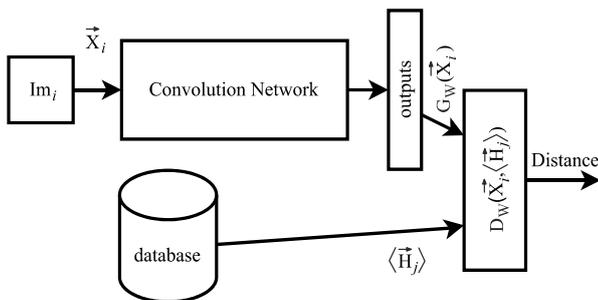


Рис. 4. Схема поиска и аутентификации пользователя, использующая БД с заранее рассчитанными хэшами

На рис. 4 представлена упрощенная функциональная структура БАС, соответствующая предложенной процедуре. Для уменьшения количества сравнений используется предварительная кластеризация векторов признаков $G_W(\vec{X})$. $\langle \vec{H}_j \rangle$ – результат кластеризации векторов ROI, полученных при регистрации j -й персоны в БАС. Процедуры предварительной обработки изображений ладоней и вычисления ROI в данной статье не рассматриваются.

4. Эксперименты

Глубокие нейросети позволяют во многих случаях достичь наилучших результатов в задачах классификации изображений, но их применение во встраиваемых приложениях ограничивается возможностями аппаратных платформ. В случае БАС одним из ограничений является максимально допустимое время аутентификации персоны.

Временная UML-диаграмма этапов аутентификации приведена на рис. 5. Видно, что время аутентификации складывается из t_1 – времени получения ROI, t_2 – времени вычисления вектора признаков (эмбедингов входных данных ИНС), t_3 – времени поис-

ка (перебора и сравнения полученных биометрических данных с данными, сохраненными в БД БАС).

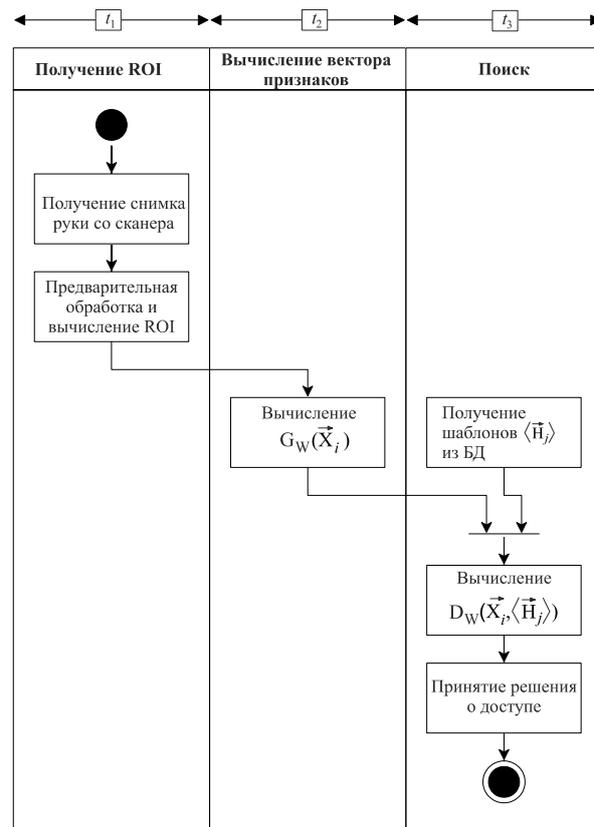


Рис. 5. Временная диаграмма работы БАС

Опуская нюансы, связанные со временем получения изображения, временем обращения к БД и другими операциями пересылки данных, можно приближенно оценить время аутентификации T_{aut} при большом количестве персон в БД:

$$T_{aut} \approx t_2 + t_3. \quad (4)$$

Время t_2 зависит от выбранного способа параметризации изображений ладоней (на основе локальных дескрипторов, хэш-функций, нейросетевых или других дескрипторов). Поскольку время вычисления метрики близости (2) зависит от размера сравниваемых векторов, то t_3 определяется не только количеством зарегистрированных в БАС персон, но и размером сохраненных шаблонов биометрических данных в БД БАС. Таким образом, задача проектирования системы осложняется не только подбором структуры сиамской нейросети, но и выбором размера выходного слоя ИНС (размера вектора признаков).

Для обучения и экспериментов с сиамскими нейросетями была собрана объединенная база, полученная из изображений датасетов CASIA-MS-PalmprintV1 collected by the Chinese Academy of Sciences' Institute of Automation (CASIA) (изображения ладоней 100 различных персон) [29] и Tongji Contactless Palmvein Dataset (изображения ладоней 300 различных персон). Датасет для обучения сформиро-

ван путем предварительной обработки и выделения регионов интереса (рис. 3) на изображениях объединенной базы ладоней (400 различных персон) [30].

В качестве аппаратной базы для встраиваемого решения БАС задана платформа Orange Pi One с процессором ARM Cortex-A7, 1.2 GHz и 512MB DDR3 SDRAM без поддержки GPU.

С учетом низкой производительности платформы Orange Pi One и ограничений на время вычислений, эмпирически определено допустимое количество и состав слоев сиамской нейросети. Одна из возможных конфигураций симметричных каналов нейросети приведена на рис. 6.

```
Conv2d(1, 16, kernel_size = (7, 7))
ReLU()
BatchNorm2d(16)
MaxPool2d(kernel_size = 3, stride = 2)
Conv2d(16, 160, kernel_size = (5, 5))
ReLU()
BatchNorm2d(160)
MaxPool2d(kernel_size = 3, stride = 2)
Dropout2d(p = 0.1)
Conv2d(160, 64, kernel_size = (3, 3))
ReLU()
BatchNorm2d(64)
Dropout2d(p = 0.1)
Linear(in_features = 25600,
out_features = 512)
ReLU(inplace = True)
Dropout2d(p = 0.3)
Linear(in_features = 512,
out_features = 512)
ReLU(inplace = True)
Linear(in_features = 512,
out_features = 128)
```

Рис. 6. Последовательность слоев симметричных каналов сиамской ИНС

Каждый канал (рис. 6) состоит из трех сверточных слоев и трех полносвязных слоев, осуществляющих сжатие карт сверточных признаков. На вход канала нейросети (рис. 6) подается изображение ROI (рис. 1) размером 105×105 пикселей с глубиной яркости 8 бит. Для вычисления потерь используется метрика Contrastive Loss (1) со значением параметра *m*, равным 2.0.

В программной реализации нейросети предусмотрен переключатель режимов работы: а) «полная сеть», используется при обучении; б) «половина нейросети», используется в процессе применения в БАС. В первом режиме активны оба симметричных канала ИНС (рис. 2), каждый из которых состоит из последовательности слоев рис.6. Во втором режиме на один из каналов подается ROI, а на второй – усредненный вектор признаков одной из зарегистрированных в базе данных БАС персон – рис. 4, рис. 5. Такой способ сравнения позволяет в процессе функционирования БАС сократить время отклика нейросети почти в два раза.

Результаты тестирования некоторых из исследованных методов классификации изображений ладоней на валидационной выборке приведены в табл. 1.

Производительность измерялась на процессоре Intel Core i5-2450M 2.5 ГГц. Аппаратная платформа Orange Pi One имеет примерно в 15 раз меньшую производительность.

Табл. 1. Сравнение эффективности методов классификации изображений ладоней

Метод	Время вычисления дескриптора (t_2)	Размер дескриптора, байт	Время поиска (t_3)	Overall Accuracy
average_hash, 8-bit	0,0004	16	0,036	73,1 %
average_hash, 16-bit	0,0006	64	0,038	92,0 %
phash, 24-bit	0,0008	144	0,044	81,8 %
average_hash, 24-bit	0,0011	144	0,046	92,8 %
average_hash, 32-bit	0,0012	256	0,051	92,9 %
vnet-160-64	0,0169	1024	0,043	93,4 %
vnet-192-64	0,0182	1024	0,048	96,9 %
vnet-208-64	0,0210	1024	0,048	96,6 %
vnet-256-64	0,0235	1024	0,052	96,8 %
vnet-256-128	0,0289	1024	0,056	97,1 %
vnet-256-256	0,0403	1024	0,068	97,8 %

В табл. 1 используются сокращения: «phash» – реализация перцептивного хэширования (Perceptual hashing) из библиотеки ImageHash; «average_hash» – реализация усредняющего хэширования (Average hashing) из этой же библиотеки; «vnet-N-M» – реализация сиамской нейросети со структурой каналов рис.6, где N – количество сверток во втором слое и M – количество сверток в третьем слое.

Сравнение показывает, что методы поиска с использованием хэширующих функций обеспечивают быстрый поиск и сравнение биометрических данных (ROI ладони) пользователей БАС, но их применение ограничивается относительно невысокими показателями качества классификации (Overall Accuracy). Усредняющее хэширование показывает несколько лучшие результаты по сравнению с перцептивным. При этом ни изменение хэширующей функции, ни изменение размера хэша не позволило в проведенных экспериментах повысить качество классификации выше некоторого предела (92,9%). Ситуация еще более ухудшается при увеличении количества пользователей, поскольку увеличение количества классов, помимо роста вероятности ошибки, влечет за собой необходимость увеличения размера вектора признаков (хэша) для исключения коллизий и, следовательно, падение скорости сравнения хэшированных биометрических данных.

При сопоставимой скорости определения пользователя БАС (табл. 1) по сравнению с методами на основе хэш-функций применение нейросетевых методов позволило повысить качество классификации по метрике Overall Accuracy до 96,9%. Нужно также отметить лучшую масштабируемость методов на основе сиамских сверточных нейросетей. Сложность структуры нейросети, количество слоев, их параметров и

размер выходного вектора признаков могут широко варьироваться в зависимости от вычислительных ресурсов аппаратной платформы и наличия графического ускорителя, что позволяет добиться более высоких показателей качества классификации.

Выводы

Нейросетевые методы часто требуют больших вычислительных ресурсов и не всегда применимы на платформах без графических ускорителей в силу временных ограничений. Однако выполненные исследования показали, что для рассмотренного в статье случая использования низкопроизводительной платформы Orange Pi One подбор параметров и режимов функционирования сямской нейросети позволяет достичь компромисса между скоростью и качеством классификации биометрических данных персон, зарегистрированных в БАС. При сопоставимом времени обработки данных качество классификации удалось повысить с 92,9% до 96,9% по метрике Overall Accuracy, что соответствует уменьшению вероятности ошибки на 56,3%.

Полученные результаты показывают, что в среднем три или четыре из ста пользователей при первом контакте с БАС будут неверно идентифицированы. Эта проблема решается вторым контактом пользователя с БАС.

Таким образом, выполненные исследования показывают возможность применения сямских сверточных нейросетей для задач биометрической аутентификации на встраиваемых платформах с ARM-архитектурой.

References

- [1] Mcconnell RK. Method of and apparatus for pattern recognition. US Patent 4,567,610 of June 28, 1986.
- [2] Freeman WT, Roth M. Orientation histograms for hand gesture recognition. Int Workshop on Automatic Face- and Gesture-Recognition, MERL-TR94-031995: 296-301.
- [3] Bosch A, Zisserman A. Pyramid histogram of oriented gradients (PHOG). 2022. Source: <<https://www.robots.ox.ac.uk/~vgg/research/caltech/phog.html>>.
- [4] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. Proc 12th Int Conf on Pattern Recognition 1994: 582-585. DOI: 10.1109/ICPR.1994.576366.
- [5] Gengjian X, Li S, Jun S, Meng W. Hybrid center-symmetric local pattern for dynamic background subtraction. IEEE Int Conf on Multimedia and Expo 2011: 1-6. DOI: 10.1109/ICME.2011.6011859.
- [6] Gupta R, Patil H, Mittal A. Robust order-based methods for feature description. IEEE Computer Society Conf on Computer Vision and Pattern Recognition 2010: 334-341. DOI: 10.1109/CVPR.2010.5540195.
- [7] Silva C, Bouwmans T, Frelicot C. An eXtended center-symmetric local binary pattern for background modeling and subtraction in videos. VISAPP 2015 – 10th Int Conf on Computer Vision Theory and Applications 2015: 395-402. DOI: 10.5220/0005266303950402.
- [8] Wu X, Sun J. An extended center-symmetric local ternary patterns for image retrieval. In Book: Lin S, Huang X, eds. Advances in computer science, environment, ecoinformatics, and education. Berlin, Heidelberg: Springer-Verlag; 2011: 359-364. DOI: 10.1007/978-3-642-23321-0_56.
- [9] Ferraz CT, Pereira O, Gonzaga A. Feature description based on center-symmetric local mapped patterns. Proc ACM Symposium on Applied Computing 2014: 39-44. DOI: 10.1145/2554850.2554895.
- [10] Narayanan V, Parsi B. Center symmetric local descriptors for image classification. Int J Nat Comput Res 2018; 7(4): 56-70. DOI: 10.4018/IJNCR.2018100104.
- [11] Datta R, Joshi D, Li J, Wang J. Image retrieval: Ideas, influences, and trends of the new age. ACM Comput Surv 2008; 40(2): 5. DOI: 10.1145/1348246.1348248.
- [12] Chatzichristofis SA, Boutalis YS. FCTH: Fuzzy color and texture histogram-a low level feature for accurate image retrieval. 9th Int Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS) 2008: 191-196. DOI: 10.1109/WIAMIS.2008.24.
- [13] Zagoris K, Chatzichristofis SA, Papamarkos N, Boutalis YS. Automatic image annotation and retrieval using the joint composite descriptor. 14th Panhellenic Conf on Informatics 2010: 143-147. DOI: 10.1109/PCI.2010.38.
- [14] David GL. Distinctive image features from scale-invariant keypoints. Int J Comput Vis 2004; 60(2): 91-110. DOI: 10.1023/B:VISI.0000029664.99615.94.
- [15] Bay H, Tuytelaars T, Van GL. SURF: Speeded up robust features. In Book: Leonardis A, Bischof H, Pinz A, eds. Computer vision -- ECCV 2006. Berlin, Heidelberg: Springer-Verlag; 2006: 404-417. DOI: 10.1007/11744023_32.
- [16] Ke Y, Sukthankar R. PCA-SIFT: a more distinctive representation for local image descriptors. 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2004: II-II. DOI: 10.1109/CVPR.2004.1315206.
- [17] Mikolajczyk K, Schmid C. A performance evaluation of local descriptors. IEEE Trans Pattern Anal Mach Intell 2005; 27(10): 1615-1630. DOI: 10.1109/TPAMI.2005.188.
- [18] Calonder M., Lepetit V, Strecha C, Fua P. BRIEF: Binary robust independent elementary features. In Book: Daniilidis K, Maragos P, Paragios N, eds. Computer vision -- ECCV 2010. Berlin, Heidelberg: Springer-Verlag; 2010: 778-792. DOI: 10.1007/978-3-642-15561-1_56.
- [19] Rublee E, Rabaud V, Konolige K, Bradski G. ORB: An efficient alternative to SIFT or SURF. 2011 Int Conf on Computer Vision 2011: 2564-2571. DOI: 10.1109/ICCV.2011.6126544.
- [20] Alcantarilla PF, Nuevo J, Bartoli A. Fast explicit diffusion for accelerated features in nonlinear scale spaces. BMVC 2013 – Electronic Proceedings of the British Machine Vision Conference 2013: 13.1-13.11. DOI: 10.5244/C.27.13.
- [21] Leutenegger S, Chli M, Siegwart RY. BRISK: Binary Robust invariant scalable keypoints. 2011 Int Conf on Computer Vision 2011: 2548-2555. DOI: 10.1109/ICCV.2011.6126542.
- [22] Tareen SAK, Saleem Z. A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK. 2018 Int Conf on Computing, Mathematics and Engineering Technologies (iCoMET) 2018: 1-10. DOI: 10.1109/ICOMET.2018.8346440.
- [23] Zauner C. Implementation and benchmarking of perceptual image hash functions. Master's Thesis, Upper Austria University of Applied 2010.

- [24] pHash. The open source perceptual hash library. 2022. Source: <<http://www.phash.org>>.
- [25] ImageHash. 2022. Source: <<https://github.com/JohannesBuchner/imagehash>>.
- [26] Sivic J, Zisserman A. Video google: A text retrieval approach to object matching in videos. Int Conf on Computer Vision 2003: 1470-1477. DOI: 10.1109/iccv.2003.1238663.
- [27] Bromley J, Guyon I, LeCun Y, Säckinger E, Shah R. Signature verification using a "Siamese" time delay neural network. Adv Neural Inf Process Syst 1993; 07(04). DOI: 10.1142/s0218001493000339.
- [28] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping. 2006 IEEE Computer Society Conf on Computer Vision and Pattern Recognition (CVPR'06) 2006: 1735-1742. DOI: 10.1109/CVPR.2006.100.
- [29] Note on CASIA Palmprint Database. 2022. Source: <<http://www.cbsr.ia.ac.cn/english/Palmprint%20Databases.asp>>.
- [30] Zhang L, Cheng Z, Shen Y, Wang D. Palmprint and palmvein recognition based on DCNN and a new large-scale contactless palmvein dataset. Symmetry 2018; 10(4): 78. DOI: 10.3390/sym10040078.

Сведения об авторах

Прозоров Дмитрий Евгеньевич, 1975 года рождения, в 1997 году окончил Вятский государственный технический университет (ныне Вятский государственный университет) по специальности «Инженер-электроник-конструктор-технолог», работает профессором на кафедре радиоэлектронных средств института математики и информационных систем Вятского государственного университета. Область научных интересов: цифровая обработка сигналов. E-mail: de_prozorov@vyatsu.ru. ORCID 0000-0002-3577-8838.

Земцов Антон Валерьевич, 1980 года рождения, окончил Вятский государственный университет по специальности «Инженер-электроник», является аспирантом направления 11.06.01 «Электроника», профиль «Радиотехнические устройства и системы», работает заместителем директора института математики и информационных систем ВятГУ. Область научных интересов: компьютерное зрение, обработка изображений, нейронные сети. E-mail: av_zemtsov@vyatsu.ru. ORSID 0000-0002-2550-653X.

ГРНТИ: 28.23.37

Поступила в редакцию 3 августа 2022 г. Окончательный вариант – 14 ноября 2022 г.

Using a lightweight Siamese neural network for generating a feature vector in a vascular authentication system

D.E. Prozorov¹, A.V. Zemtsov¹

¹ Vyatka State University, 610000, Kirov, Russia, Moskovskaya 36

Abstract

The article analyzes the possibility of using a Siamese convolutional neural network to solve the problem of vascular authentication on an embedded hardware platform with limited computing resources (Orange Pi One). The authors give a brief review of modern methods for calculating image feature vectors used in the tasks of classifying, comparing or searching for images by content: based on variational series (histograms), local descriptors, singular point descriptors, descriptors based on hash functions, neural network descriptors. They suggest using the architecture of a biometric authentication system (BAS) based on images of palms in the visible and near-IR spectra based on a Siamese convolutional neural network. The developed software solution allows using the Siamese neural network in the "full network" (both symmetrical channels of the neural network are used) and "half of the neural network" (only one channel is used) modes to reduce the time for comparing biometric data vectors - images of the palms of registered BAS users. The authors demonstrate advantages of the neural network features: universality, scalability and competitiveness, including on embedded hardware and software solutions with limited computing resources without graphics accelerators. The studies have shown that using the Siamese neural network, the "overall accuracy" of palm image classification can be improved from 0.929 to 0.968 when compared with the image vectorization method based on a perceptual hash, while showing a comparable authentication time for individuals registered in BAS. In the experiments, the authors use a database of 2,000 images for 400 people.

Keywords: biometric authentication, image processing, image descriptors, artificial neural network, Siamese neural network.

Citation: Prozorov DE, Zemtsov AV. Using a lightweight Siamese neural network for generating a feature vector in a vascular authentication system. *Computer Optics* 2023; 47(3): 433-441. DOI: 10.18287/2412-6179-CO-1204.

Authors' information

Dmitry Evgenievich Prozorov (b. 1975) graduated from Vyatka State Technical University (now Vyatka State University) in 1997 with a degree in Electronics-Design-Technology, works as a professor of Radioelectronic Means department of the Institute of Mathematics and Information Systems of Vyatka State University. Research interests are digital signal processing. E-mail: de_prozorov@vyatsu.ru.

Anton Valerievich Zemtsov (b. 1980) graduated from Vyatka State University with a degree in Electronics Engineering, a postgraduate student of the direction 11.06.01. Electronics; profile "Radio Engineering Devices and Systems", works as Deputy Director of the Institute of Mathematics and Information Systems of Vyatka State University. Research interests are computer vision, image processing, neural networks. E-mail: av_zemtsov@vyatsu.ru.

Received August 3, 2022. The final version – November 14, 2022.
