Интеллектуальные системы управления, анализ данных

© 2023 г. Х. ЧЕН, PhD (eric.hf.chen@hotman.com)
(Чжэцзян Шурен университет, Ханчжоу),
С.А. ИГНАТЬЕВА (s.ignatieva@psu.by),
Р.П. БОГУШ, д-р техн. наук (r.bogush@psu.by)
(Полоцкий государственный университет
имени Евфросинии Полоцкой, Новополоцк),
С.В. АБЛАМЕЙКО, д-р техн. наук (ablameyko@bsu.by)
(Белорусский государственный университет, Минск)

ПОВТОРНАЯ ИДЕНТИФИКАЦИЯ ЛЮДЕЙ В СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ С ИСПОЛЬЗОВАНИЕМ ГЛУБОКОГО ОБУЧЕНИЯ: АНАЛИЗ СУЩЕСТВУЮЩИХ МЕТОДОВ

Статья посвящена многостороннему анализу повторной идентификации людей в системах видеонаблюдения и современных методов ее решения с использованием глубокого обучения. Рассматриваются общие принципы и применение сверточных нейронных сетей для этой задачи. Предложена классификация систем реидентификации. Приведен анализ существующих наборов данных для обучения глубоких нейронных архитектур, описаны подходы для увеличения количества изображений в базах данных. Рассматриваются подходы к формированию признаков изображений людей. Представлен анализ основных применяемых для реидентификации моделей архитектур сверточных нейронных сетей, их модификаций, а также методов обучения. Анализируется эффективность повторной идентификации на разных наборах данных, приведены результаты исследований по оценке эффективности существующих подходов в различных метриках.

Ключевые слова: реидентификация, видеоданные, сверточные нейронные сети, метрики оценки точности, дескрипторы изображений.

DOI: 10.31857/S0005231023050057, **EDN:** AHHWFO

1. Введение

Широкое внедрение систем видеонаблюдения позволяет решать множество практических задач, в том числе и повышения уровня общественной безопасности. Так, важным и актуальным является определение присутствия заданного человека по его изображениям на видеоданных в другом месте или

в разное время в пространственно-распределенных системах видеонаблюдения. Такая задача называется повторной идентификацией или реидентификацией человека. Для ее решения необходимо выявить отличительные признаки и путем выполнения запроса к базе данных сравнить их с признаками из имеющейся выборки изображений множества людей (галереи). Причем состав набора признаков в значительной мере определяет эффективность реидентификации. Поиск и выделение наиболее отличительных особенностей объектов на изображениях, в том числе и людей, не формализованы. Следовательно, используется эмпирический подход, который в большинстве случаев является долгим и трудоемким процессом. Для реидентификации людей из-за неоднозначности внешнего вида с разных ракурсов, вариаций освещения, различных разрешений камер, окклюзий для этого требуются нерационально большие вычислительные затраты. Поэтому долгое время для повторной идентификации людей значимые результаты не достигались. Совершенствование средств вычислительной техники и открытия в области глубокого обучения, в частности развитие сверточных нейронных сетей (СНС), позволили автоматизировать процесс извлечения признаков изображений людей и обеспечить значительное увеличение точности реидентификации. Однако несмотря на то, что данной задачей с применением методов глубокого обучения занимаются многие ученые и инженеры в мире, она не решена полностью, и при разработке системы повторной идентификации по-прежнему приходится сталкиваться с большим числом проблем, а широкое разнообразие областей применения повторной идентификации, таких как пропускные системы на режимных предприятиях, поиск пропавших людей или правонарушителей, сбор статистической информации о посещении людьми торговых центров и других социальных объектов, приводят к существованию большого числа подходов и алгоритмов для ее решения, которые имеют разные качественные характеристики.

2. Организация и оценка эффективности повторной идентификации людей в распределенных системах видеонаблюдения

2.1. Обобщенная схема системы повторной идентификации

Пространственно-распределенная система видеонаблюдения состоит из территориально разнесенных IP камер и организована, как правило, на основе единого центра обработки данных. На рис. 1 показана упрощенная структура повторной идентификации в такой системе, которая включает три IP видеокамеры C_1, C_2, C_3. На каждом кадре F^k, k — номер видеокамеры, с помощью детектора выполняется обнаружение всех людей, попадающих в поле зрения камер, и формирование ограничительных рамок для них, которые описывают прямоугольником обнаруженные фигуры. Изображения людей I_i, где $i = 1, \ldots, N_{img}$, N_{img} — общее количество изображений, размещаются в галерее. Для каждого из них с помощью СНС определяются векто-

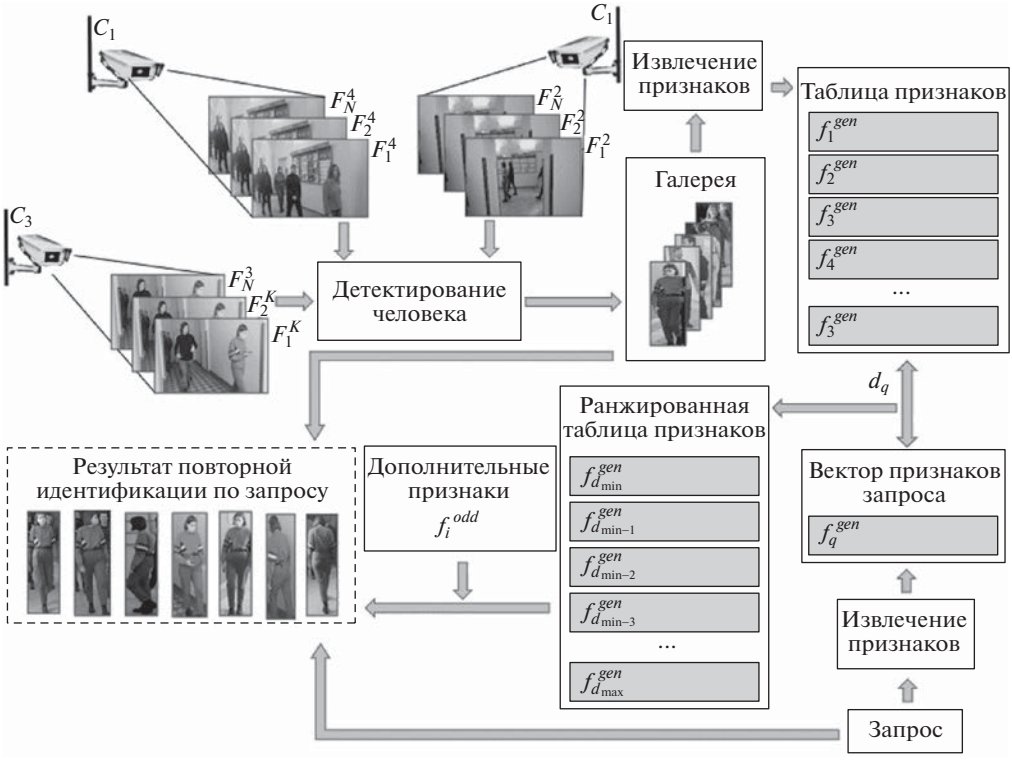

Рис. 1. Общая схема системы повторной идентификации.

ры f_i^{gen} (CHC дескрипторы), формирующие общее пространство CHC признаков $\chi_{Ii} = \{f_i^{gen}\}$, которое представляется в виде таблицы, причем каждая строка является CHC дескриптором f_i^{gen} для одного изображения.

Для описания человека при редентификации используется составной вектор признаков, который может быть представлен как

$$P_{ID} = (p_n^{ID}, f_i^{gen}, f_i^{add}), \tag{1}$$

где p_n^{ID} — идентификатор (метка) человека; n — количество возможных идентификаторов, равное общему числу уникальных людей; f_i^{gen} — вектор CHC признаков для i-го изображения человека, который может включать CHC признаки, разделяемые на глобальные признаки f_i^{global}, характеризующие изображение в целом, и локальные $f_{i,j}^{local}$, получаемые при разделении изображение на j частей; f_i^{add} — дополнительные признаки, которые могут содержать информацию, позволяющую улучшить эффективность системы реидентификации, например идентификатор камеры C_{ID}, номер кадра с k-й видеокамеры F_m^k или др. [1].

При поступлении запроса для повторной идентификации человека вычисляется его вектор признаков f_q^{gen}, который используется для нахождения расстояния d_q, определяющего степень подобия между данным запросом и

дескрипторами изображений галереи. С использованием найденных расстояний выполняется ранжирование в таблице χ_{I_i} от $d_{\min}$ до $d_{\max}$. С учетом дополнительных признаков исключаются изображения, которые по каким-либо критериям позволяют предполагать, что несмотря на схожесть визуальных признаков, изображение-кандидат не соответствует искомому человеку. Например, если на изображениях с двух неперекрывающихся камер в одно и то же время находится объект интереса со схожими визуальными признаками, то можно однозначно утверждать, что это разные люди, так как один и тот же человек не может присутствовать в двух местах одновременно. После исключения всех неподходящих кандидатов в качестве результата повторной идентификации отображаются изображения людей, f_i^{gen} которых находились вверху списка ранжированной таблицы. Первый человек из этого списка принимается за результат повторной идентификации как наиболее схожий с запросом.

2.2. Классификация систем повторной идентификации

Широкая область применения систем повторной идентификации человека обуславливает существование большого количества алгоритмов и подходов для решения задачи, и, соответственно, различные способы классификации таких систем (рис. 2). Так, по взаимодействию с внешней средой можно выделить системы повторной идентификации закрытые (Close-world), использующие готовые наборы данных для обучения и тестирования, и открытые (Open-world), в которых галерея изображений постоянно пополняется новыми кадрами [2]. Закрытые системы обычно применяются в исследовательских целях и набор данных состоит из ограниченного количества видеопоследовательностей или изображений, полученных с нескольких камер видеонаблюдения. Данные в таких наборах аннотированы и подготовлены заранее, запрос присутствует в галерее. В открытых системах используется набор данных,

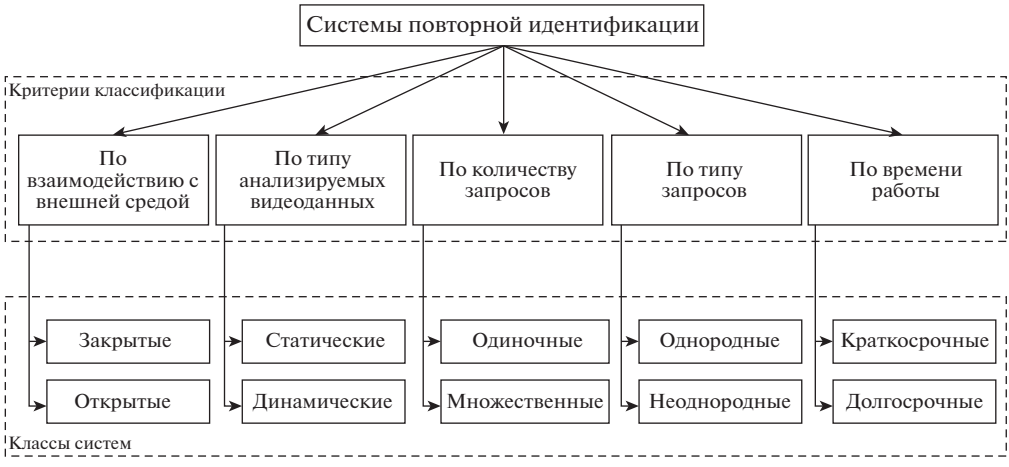

Рис. 2. Классификация систем повторной идентификации.

который изменяется с течением времени, по мере поступления новых записей с камер наблюдения ограничительные рамки необходимо генерировать в режиме реального времени. Полученные новые изображения требуется аннотировать для обучения, т.е. формировать псевдо-метки (pseudo-label) для возможности тренировки СНС при видеонаблюдении. Организация таких систем намного сложнее, они требуют высокопроизводительной аппаратурной части, но наиболее приближены к реальным условиям.

В зависимости от типа анализируемых видеоданных системы повторной идентификации можно разделить на статические (image-based), которые обрабатывают отдельные кадры через некоторые интервалы времени, и динамические (video-based), когда рассматривается последовательность кадров из видео [3]. В динамических системах признаки формируются не только на основе анализа пространственной области, но и учитывают временную составляющую о человеке, например информацию о походке, направлении движения и другие дополнительные признаки.

В зависимости от количества запросов [4] системы реидентификации можно разделить на одиночную повторную идентификацию (для одного человека) и множественную (для всех людей, попавших в поле зрения камер). В первом случае в наборе данных требуется найти человека по запросу, и повторная идентификация сводится к задаче поиска или проверке, присутствует ли искомый человек в галерее. Во втором — для каждого человека устанавливается уникальный идентификатор и определяется, на каких кадрах каждый из этих людей встречается снова, и эта задача сводится к классификации [5].

По типу запросов системы повторной идентификации можно разделить на однородные (single-modality) и неоднородные (cross-modality) [2]. При использовании однородных данных в качестве запросов используются изображения или видео, полученные с камер видеонаблюдения видимого диапазона. Если в качестве запроса используется текстовое описание искомого человека, изображение с инфракрасной камеры, рисунок или эскиз, то такие системы будут называться неоднородными.

По времени работы системы выделяют краткосрочную повторную идентификацию и долгосрочную [6]. Так, если каждый человек на изображениях в наборе данных находится в одной и той же одежде, изменения внешности незначительны и обусловлены только возможным изменением наличия аксессуаров или вещей в руках, съемка осуществлялась в течение ограниченного интервала времени, за которое человек не мог значительно изменить образ, то такая система будет краткосрочной. Долгосрочная повторная идентификация направлена на способность повторно идентифицировать людей, даже если прошло уже значительное количество времени, за которое человек мог изменить внешний вид [7].

Любая из рассмотренных выше систем может столкнуться с проблемой смещения домена (domain shift), когда обучение и тестирование осуществляются на данных из разных доменов. Под доменом понимают комплект изоб-

ражений, которые были получены в одинаковых условиях в одной системе видеонаблюдения. На каждое изображение в наборе данных оказывает влияние совокупность факторов, включающих разрешение камер, фон, условия освещения и даже внешний вид людей, т.е. статистически европейцы будут иметь отличный вид от азиатов, летняя одежда от зимней и т.д. Система, обученная на наборе данных, полученном с внутренних камер видеонаблюдения, может иметь крайне низкую эффективность на тестовой выборке, состоящей из изображений людей с наружных камер видеонаблюдения. Алгоритмы, направленные на решение этой проблемы, называются «междоменной реидентификацией» (Cross-domain ReID) и реализуют задачу адаптации (или переносимости) домена.

2.3. Метрики оценки точности

Одним из важнейших вопросов для оценки результатов повторной идентификации является выбор метрик, позволяющих дать численную оценку эффективности алгоритма и сравнить результаты для разных подходов реидентификации. Наиболее распространенными является группа метрик RankN, включающая Rank1, Rank5, Rank10, и mAP. Группа метрик RankN характеризует качество ранжирования и показывает процент числа запросов, для которых верный выданный результат был среди первых N полученных результатов. Соответственно, метрика Rank1 показывает процент запросов, для которых идентификатор первого изображения-кандидата совпадает с идентификатором запроса. Если $\mathrm{N} = 5$, то Rank5 показывает процент запросов, для которых среди первых пяти выданных изображений-кандидатов было верное решение, соответственно для Rank10 учитываются первые десять изображений-кандидатов. Для вычисления RankN определяется отношение суммы числа запросов, для которых верное решение было найдено среди первых выданных результатов, к общему числу запросов Q:

$$\mathrm{RankN} = \frac{\sum K_{i,\mathrm{N}}}{Q}, \tag{2}$$

где i — номер запроса; $K_{i,\mathrm{N}}$ — i-й запрос, для которого верное решение было найдено среди первых N выданных результатов.

Метрика mAP является оценкой точности алгоритма повторной идентификации, отражающей среднее значение средних точностей для всех запросов, и рассчитывается по формуле

$$\mathrm{mAP} = \frac{1}{Q}\sum_{i=1}^{Q} AP_i, \tag{3}$$

где AP — средняя точность, определяемая как площадь под кривой *precision–recall*, где $precision = \frac{TP}{TP+FP}$ — точность, TP — количество верных предсказаний запросов; FP — количество ложных положительных предсказаний

запросов; $recall = \frac{TP}{TP+FN}$ — чувствительность; FN — количество ложных отрицательных предсказаний запросов.

В системах повторной идентификации приоритетно, чтобы верные предсказания находились в начале ранжированного списка и имели как можно меньше ложных предсказаний. Следует отметить, что метрики RankN и mAP не отражают сложность поиска правильно идентифицированных изображений людей для поступающего запроса. Кроме этого, при одинаковых показателях Rank точность AP может отличаться. Для учета поиска наиболее сложных правильных предсказаний в [8] используется метрика mINP (mean Inverse Negative Penalty), предложенная в [2], которая позволяет исключить доминирование легких совпадений, влияющих на метрики Rank и mAP. Для ее вычисления вводятся дополнительные метрики: NP (Negative Penalty) — отрицательный штраф, назначаемый за неверные предсказания для i-го запроса и уменьшающий вероятность правильной реидентификации при неправильном нахождении самого сложного совпадения; INP (Inverse Negative Penalty) — обратная величина для NP, рост которой свидетельствует о повышении эффективности системы. При этом mINP характеризует среднее значение INP для всех запросов и вычисляется как

$$\text{mINP} = \frac{1}{Q}\sum_i (1 - \text{NP}_i) = \frac{1}{Q}\sum_i \left(1 - \frac{R_i^{hard} - |G_i|}{R_i^{hard}}\right) = \frac{1}{Q}\sum_i \frac{|G_i|}{R_i^{hard}}, \tag{4}$$

где $\text{NP}_i = \frac{R_i^{hard} - |G_i|}{R_i^{hard}}$ — отрицательный штраф; R_i^{hard} — позиция самого сложного верного предсказания; $|G_i|$ — общее количество верных предсказаний для запроса.

На рис. 3 показан пример, когда в галерее для каждого запроса есть только три верных изображения (True). В первых двух ранжированных списках на рис. 3 при одинаковом значении Rank1 метрики AP различны: $\text{AP}_1 = 0{,}77$ (см. рис. 3,*а*), $AP_2 = 0{,}63$ (см. рис. 3,*б*). Это связано с тем, что в начале первого ранжированного списка имеются два верных совпадения, а во втором — только одно. При этом ближайшее верное совпадение занимает пятую позицию. Если сравнивать списки на рис. 3,*б* и 3,*в*, то очевидно, что в третьем ранжированном списке $\text{AP}_3 = 0{,}64$, т.е. больше, чем во втором, но при этом Rank1 в этом примере равен нулю. Это так же объясняется тем, что все возможные правильные ответы были получены вверху ранжированной таблицы (на второй, третьей и четвертой позиции), за исключением первого, неверного, предсказания. Предпочтительнее, чтобы все верно идентифицированные изображения людей были получены как можно раньше, однако при оценке системы метрики AP и Rank не позволяют это определить с максимальной точностью.

Анализ рис. 3,*в* показывает, чтобы иметь все возможные верные ответы, необходимо получить только четыре первых изображения-кандидата, и соответственно отрицательный штраф будет равен $\text{NP} = 0{,}25$, который минимальный для примеров на рис. 3. На рис. 3,*б* самое сложное предсказание

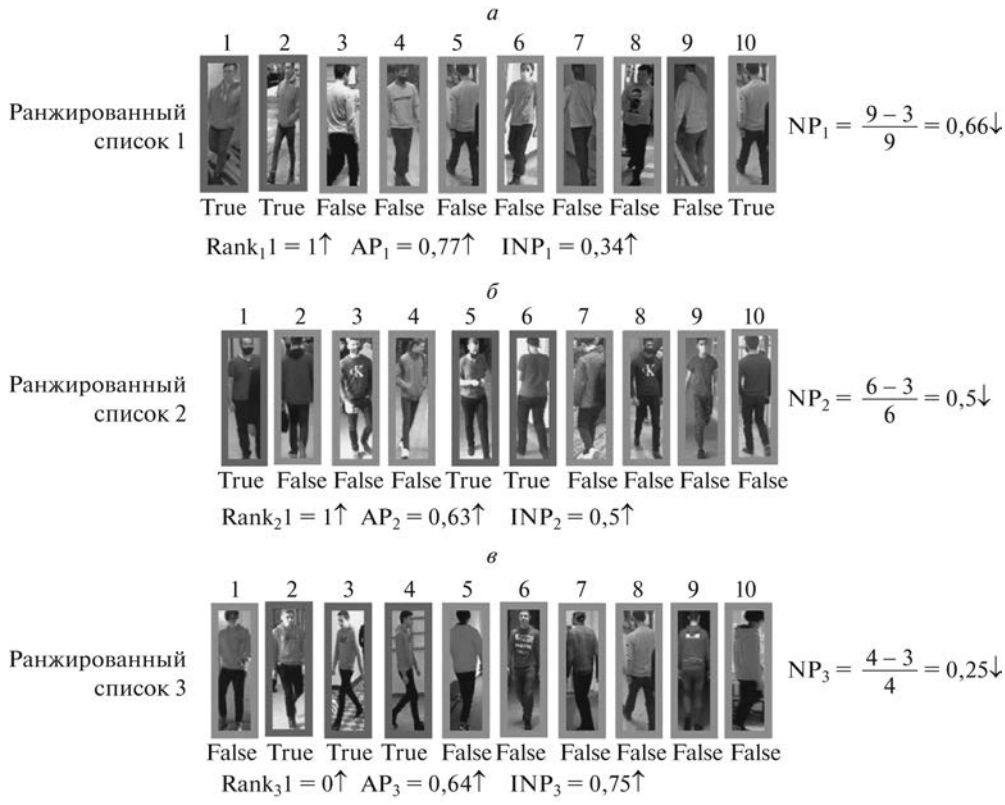

Рис. 3. Различие в метриках Rank, AP, NP и INP в зависимости от позиции истинных и ложных предсказаний.

соответствует шестой позиции в ранжированной таблице, на рис. 3,*а* неправильное обнаружение человека характерно для девятой позиции. Поэтому для примеров на рис. 3,*а* и 3,*б* увеличиваются значения NP, соответственно уменьшается INP. Таким образом, метрика INP позволяет оценить влияние сложности поиска всех верных совпадений. Чем больше это значение, тем лучше система выполняет поиск всех людей с одинаковым идентификатором. Соответственно следует стремиться к снижению NP и уменьшать число позиций от начала списка ранжирования до самого сложного, который может быть неправильно идентифицирован при поиске изображения.

3. Наборы и подготовка данных для обучения СНС

Использование СНС для извлечения признаков приводит к необходимости обучения используемой модели глубокой нейронной сети. Для этой цели обычно применяется аннотированный набор данных, который содержит уникальный идентификатор для каждого отдельного человека $S = \{(I_i, p_1^{ID}), \ldots, (I_m, p_n^{ID})\}$, где I_i — изображение, $1 \leqslant i \leqslant m$, m — количество изображений, p_n^{ID} — идентификатор человека. Часто изображения сопровож-

даются информацией о номере камеры, с которой они были получены, номере кадра в видеопоследовательности. В аннотированном наборе данных для эффективной работы системы необходимо извлекать такой вектор признаков $f^{gen}(I_i)$ чтобы во всем пространстве признаков χ_{I_i} расстояние между ними для одинаковых идентификаторов было меньше, чем для людей с разными метками, т.е. следует стремиться к уменьшению ошибки E предсказания идентичности в S

$$\min E(I_i, p_n^{ID}) \in [p_n^{ID} - g(f^{gen}(I_i))], \tag{5}$$

где g — классификатор. Качество извлеченных признаков зависит от распределения и разнообразия данных в S [9].

При тренировке СНС для улучшения точности повторной идентификации рекомендуется подобирать наиболее оптимальные гиперпараметры, такие как скорость обучения, размер пакета, количество эпох; использовать увеличение обучающей выборки, аугментацию данных, найти наиболее эффективную функцию потерь, архитектуру СНС или рассматривать изображение не целиком, а разделяя его на фрагменты.

Для уже обученной модели улучшение работы алгоритма можно достигнуть, подбирая наиболее эффективный способ ранжирования таблицы признаков, использовать повторное ранжирование, учитывать дополнительную информацию о времени и месте съемки, атрибутах. Под атрибутами понимают семантическую информацию о человеке, имеющую значение для его идентификации. К ним относятся цвет и вид одежды, длина волос человека, наличие и особенности сумки, рюкзака, очков и других значимых деталей.

3.1. Анализ наборов данных

На точность повторной идентификации существенное влияние оказывают размер и состав обучающей выборки. Однако алгоритм для реидентификации в значительной мере определяет требования к набору данных. Формирование банка изображений для обучения и тестирования представлят трудоемкий и длительный процесс. Кроме этого существует проблема сдвига домена [10, 11], когда наблюдается значительное снижение точности повторной идентификации при использовании системы в условиях, стилистически отличающихся от обучающей выборки. Частичным решением проблемы является объединение разных наборов данных, что рассматривается в [12, 13], в том числе и для необходимого домена [12, 14].

При использовании существующих наборов данных для обучения СНС, кроме проблемы сдвига домена, приходится сталкиваться с проблемой защиты персональных данных. Некоторые базы изображений являются закрытыми, в них авторы предоставляют для исследований только извлеченные признаки [15]. Другие наборы данных можно использовать с ограничениями [16–18], т.е. при публикации исследований авторы просят соблюдать конфиденциальность студентов, изображения которых использовались при со-

здании, и распространение этих баз изображений возможно только при согласовании с авторами. Для некоторых наборов данных ограничивается возможность их использования. Например, MSMT17 [19] в настоящее время не доступен в публичном доступе, а DukeMTMC-ReID [20] был отозван и его использование не рекомендуется [21].

Существующие наборы изображений отличаются количеством сцен съемки и разных людей, а также числом изображений для каждого отдельного человека. Такие базы данных могут содержать отдельные кадры целиком, например PRW [22] и CUHK-SYSU [23], или вырезанные с этих кадров прямоугольные фрагменты на основе ограничительных рамок, содержащие только изображение человека. В некоторых наборах данных включены комплекты ограничительных рамок, полученных с нескольких последовательно идущих кадров, которые называются треклетами (tracklets), например MARS [24], LPW [25]. Также могут содержаться ограничительные рамки, полученные с отдельных кадров, взятых с некоторым интервалом по времени, например Market-1501 [26], CUHN01 [16], CUHN02 [17], CUHN03 [18], VIPer [27] и др.

Изображения для наборов данных, как правило, получены при различных условиях съемки вне помещений (Market-1501 [26], LPW [25], PRID [28]) или в помещениях (QMUL iLIDS [29], Airport [30]). При формировании базы изображений PolReID [31] использовалось 856 сцен сьемки внутреннего и наружного наблюдения. В наборе данных CUHN01 изображения для каждого человека получены с двух камер, области обзора которых не пересекаются. В CUHN02 используется пять таких пар видеокамер, а в CUHN03 изображения формируются с шести видеокамер, но для каждого человека предоставляются ограничительные рамки только с двух. Набор данных VIPeR был сформирован на основе изображений, полученных с двух видеокамер наружного видеонаблюдения, и для каждого человека представлено всего по одному изображению с каждой из них. При формировании LPW использовалось три разных локации, и на первой локации было установлено три видеокамеры, на двух других по четыре. Наборы данных PRW, Market-1501 и MARS были получены в одном и том же месте возле супермаркета в университете Циньхуа с шести видеокамер и отличаются только способом представления данных: кадры целиком, ограничительные рамки с изображением человека, треклеты соответственно.

Для обучения и тестирования неоднородных систем повторной идентификации применяются специальные наборы данных, использующие в качестве запроса текст (CUHK-PEDES [32], ICFG-PEDES [33]), изображение низкого разрешения (LR-PRID [34], LR-VIPeR [35]), изображение с инфракрасной камеры (SYSU-MM01 [36], RegDB [37]) или эскиз (PKU-Sketch [38]).

Набор данных CUHK-PEDES [32] объединяет пять существующих, таких как CUHK03 [18], Market-1501 [26], SSM [39], VIPeR [27] и CUHK01 [16], и каждое изображение аннотируется двумя текстовыми описаниями на английском языке. Текстовое описание состоит в среднем из 23,5 слов и содержит

информацию о внешнем виде человека, его действиях, позах. Другим набором данных для неоднородных систем повторной идентификации является ICFG-PEDES [33], который содержит в среднем 37,2 слов с более детальным описанием внешности, чем CUHK-PEDES, и сформирован на основе MSMT17 [19].

Наборы данных LR-PRID [34], LR-VIPeR [35] получены с использованием PRID [28] и VIPeR [27] соответственно, и для каждого человека имеется пара изображений, одно из которых с низким разрешением, а другое с большим, что позволяет их применять для систем повторной идентификации с видеокамерами разного разрешения.

SYSU-MM01 [36] был получен с двух инфракрасных и четырех RGB-камер, состоит из 15 712 инфракрасных изображений и 22 559 цветных для 491 человека. Набор RegDB [37] содержит по 10 цветных изображений, снятых днем, и 10 тепловых изображений с ночной ИК-камеры для 412 человек, что определяет возможность их использования в неоднородных системах повторной идентификации с инфракрасными и RGB-видеокамерами.

В [38] предлагается набор данных для двухсот человек, включающий по два изображения с разных камер и эскиз для каждого. Для создания эскизов были привлечены волонтеры, которые описывали внешность людей пяти разным художникам для обучения открытой (open-world) неоднородной (cross-modality) системы повторной идентификации. В случае отсутствия фотографии человека используется эскиз, нарисованный по описанию.

Еще одним набором данных для открытых (open-world) систем повторной идентификации является MPR Drone [40], который отличается тем, что для получения изображений используется одна видеокамера летающего дрона. Весь набор состоит из двух частей, первая часть размечена для 113 610 обнаруженных ограничительных рамок, а вторая содержит необработанные кадры для первой части.

В [41] представлен большой немаркированный набор данных LUPerson, который включает более четырех миллионов изображений для двухсот тысяч человек и может использоваться для неконтролируемого обучения систем повторной идентификации. Он сформирован с использованием видеоданных с более чем семидесяти тысяч уличных видео из различных городов.

В табл. 1 приведены сравнительные характеристики рассмотренных наборов данных.

В связи с тем, что при создании набора данных необходимо явное согласие всех участников, некоторые исследователи для формирования обучающей выборки применяют сгенерированные изображения. В [42] предлагается синтетический набор данных для повторной идентификации людей MOTSynth, для создания которого использовались видеопоследовательности из игры Grand Theft Auto V (GTA-V), имитирующей город с жителями в трехмерном пространстве. Авторы вручную разметили точки обзора камеры, спланировали маршруты и перемещения пешеходов, установили параметры,

Таблица 1. Сравнительная таблица наборов данных для повторной идентификации

Набор данных	Количество камер	Количество человек	Количество ограничительных рамок	Размер изображения
PRW [22]	6	932	34 304	Различный
CUHK-SYSU [23]	6	8432	96 143	–
MARS [24]	6	1261	1 191 003	256×128
LPW [25]	3, 4, 4	2731	592 438	256×128
Market-1501 [26]	6	1501	32 217	128×64
CUHN01 [16]	2	971	3884	160×60
CUHN02 [17]	10 (5 пар)	1816	7264	160×60
CUHN03 [18]	6	1360	13 164	Различный
MSMT17 [19]	15	4101	126 441	Различный
VIPeR [27]	2	632	1264	128×48
PRID [28]	2	934	24 541	128×64
QMUL iLIDS [29]	2	119	476	Различный
Airport [30]	6	9651	39 902	128×64
PolReID [31]	856	657	52 035	Различный
CUHK-PEDES [32]	–	13 003	80 412	–
ICFG-PEDES [33]	–	4102	52 522	–
LR-PRID [34]	2	100	200	–
LR-VIPeR [35]	2	632	1264	128×48 и 64×24
SYSU-MM01 [36]	6	491	38 271	–
RegDB [37]	2	412	8240	–
PKU-Sketch [38]	2	200	400	–

связанные с поведением людей, характерным для людных мест. Анализировалось 597 различных моделей пешеходов, для которых случайным образом менялась одежда, рюкзаки, сумки, маски, прически и бороды. Это позволило получить более 9519 уникальных пешеходов. Приведенные авторами результаты показывают, что обучение на синтетическом наборе позволяет повысить точность реидентификации на 6,9 % в метрике mAP по сравнению с использованием для обучения Market-1501 [26] и на 2,5% в метрике mAP при обучении на объединенном наборе данных из Market-1501[26] и CUHK03 [18].

В [9] рассматривается алгоритм генерации синтетических изображений для повышения устойчивости системы к смене домена. Для создания трехмерных реалистичных изображений людей применяется MakeHuman [43], а для моделирования видеонаблюдения платформа — Unreal Engine 4 (UE4) [44] с возможностью регулирования условий съемки (ночная, в помещении, на

улице), количества окклюзий людей, скорости ходьбы. Используется большое число деталей внешности, таких как маски, очки, наушники, головные уборы. На полученных изображениях людей присутствуют реальные фрагменты одежды, что отличает данный подход от существующих. При генерации намеренно добавляются люди с похожей внешностью и небольшими отличительными особенностями. Представлены результаты исследования в [9], которые показывают, что применение данного набора позволяет получить большую точность Rank1 при междоменном тестировании с использованием MSMT17, по сравнению с применением других синтетических баз изображений, таких как SOMAset [45], SyRI [46], PersonX [47], RandPerson [48]. Результаты подтверждаются при тестировании на Market-1501 и DukeMTMC-ReID.

В [49] предлагается синтетический набор данных ClonedPerson, содержащий 3D-изображения людей, при этом одежда всех сгенерированных персонажей клонируется с реальных изображений, что позволяет усилить сходство между виртуальным человеком и его прототипом. Всего набор данных включает 887 766 изображений для 5621 человека. Для генерации изображений использовалась платформа Unity3D [50], как и для RandPerson [48]. Полученный таким образом набор данных используется для обучения СНС и позволяет достигнуть лучших результатов при тестировании на изображениях из другого домена в метрике mAP на CUHK03[18], Market-1501 [26], MSMT17 [19] по сравнению с применением для обучения RandPerson [48] и UnrealPerson [9]. Следует отметить, что существенным преимуществом синтетических наборов данных является автоматическая генерация аннотаций.

3.2. Аугментация обучающей выборки

Увеличение объема обучающей выборки за счет модификации имеющихся в ней изображений называют аугментацией. Традиционными подходами для этого являются различные преобразования изображений, такие как поворот, отражение, изменение размера, контраста, яркости, вариации цветовой составляющей, размытие. Для повышения устойчивости к окклюзиям применяется метод «случайного стирания» [51]. При этом прямоугольный фрагмент изображения, размер и форма которого выбирается произвольным образом, заполняется нулевыми или случайными значениями (рис. 4). Тестирование данного метода аугментации для реидентификации осуществлялось на наборах Market-1501, DukeMTMC-ReID и CUHK03. Результаты исследований показали, что в некоторых случаях, например при тестировании на CUHK03, такой способ позволяет повысить точность почти на 9% в метрике Rank1 и более чем на 6% в метрике mAP. При использовании Market-1501 и DukeMTMC-ReID для разных алгоритмов точность в метриках Rank1 и mAP была увеличена на 1–4%.

Следует отметить, что в алгоритмах повторной идентификации аугментация данных используется для увеличения обучающей выборки путем случайного выбора изображения для какого-либо преобразования, но механизм этого влияния не рассматривается, т.е. как факт принимается то, что это

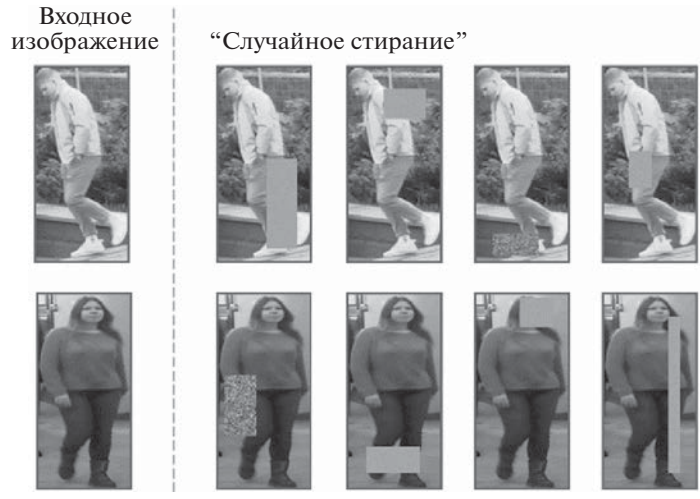

Рис. 4. Примеры применения стирания фрагмента изображения для аугментации данных.

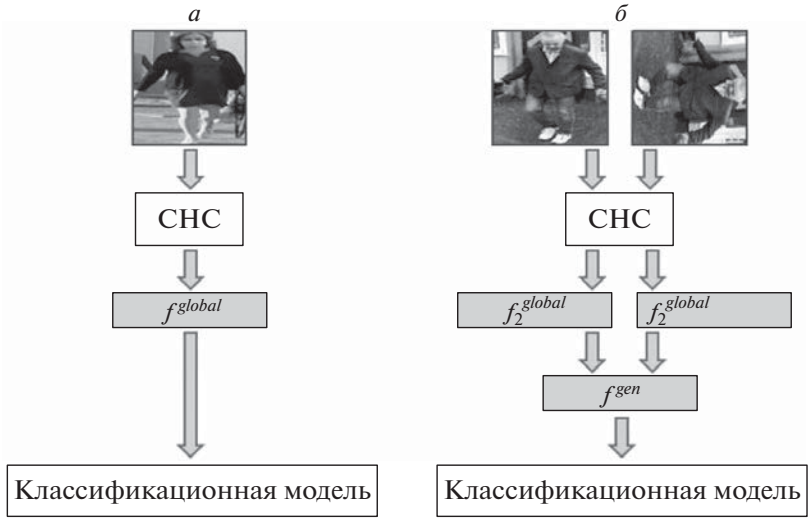

Рис. 5. Принцип извлечения признаков: *а* – базовый, *б* – с учетом поворота изображений [52].

позитивно влияет на точность работы обученной модели за счет улучшения обобщающей способности сети. В [52] используются повороты для увеличения количества изображений, но при этом тренировка СНС осуществляется одновременно как для исходного изображения, так и для преобразованных, и оцениваются потери, возникающие при повороте, что позволяет минимизировать среднеквадратичную ошибку между векторами признаков для соответствующей пары изображений. На рис. 5 представлено сравнение базового алгоритма аугментации данных при обучении извлечению признаков и предложенно-

го в [52]. Базовый алгоритм предполагает, что используется один поворот для случайного образца за один проход по сети при обучении, а рассмотренный в [52] прдполагает, что каждое изображение поворачивается на случайный угол и подается на вход сети одновременно с исходным. С помощью СНС из пары изображений извлекаются признаки, которые затем усредняются. При сравнении с другими алгоритмами повторной идентификации предложенный позволил повысить точность в метрике mAP: более чем на 5% для Market-1501, более 10% для DukeMTMC-reID и более 20% для MSMT-17 по сравнению с базовым алгоритмом (рис. 5,*а*). При этом для MSMT-17 достигнуто максимальное значение точности повторной идентификации в метриках mAP = 81,3 и Rank1 = 87,5 на момент публикации работы [52]. В [53] предложен алгоритм, повышающий значения метрик mAP = 84,4 и Rank1 = 89,9 для набора данных MSMT-17.

Более сложным методом аугментации данных является применение генеративно-состязательных сетей (Generative Adversarial network — GAN), которые используются для генерации изображений, близких к естественным, на основе уже имеющихся данных. Генеративно-состязательная сеть представляет собой алгоритм машинного обучения, в основе которого лежит комбинация двух нейронных сетей. Одна из них генерирует изображения, а другая пытается определить, могут ли они быть онесены к подлинным. Применительно к реидентификации использование GAN может быть направлено на улучшение способности извлечения эффективных признаков [54] или на решение задач со смещением доменов [55].

В [54] рассматривается проблема, характерная для повторной идентификации в реальных условиях, когда возможно присутствие различных факторов, ухудшающих качество изображений, полученных с камер видеонаблюдения. Например, если в момент наблюдения идет дождь, то система, обученная на данных, полученных при других условиях, не сможет с высокой точностью интерпретировать извлеченные дескрипторы. В подобных случаях существует высокая вероятность того, что большое число сформированных признаков будет учитывать сходства не людей, а факторов, ухудшающих качество изображения. Для решения этой проблемы необходимо изучить признаки различных явлений, снижающих качество изображений. Однако в реальных условиях сложно получить аннотации для описания подобных возмущающих воздействий, а в обучающей выборке может не быть эталонных примеров. Для извлечения робастных к ухудшающим факторам изображений признаков авторы используют GAN для синтезирования изображений с заранее известной степенью искажения.

В [55] GAN применяется для аугментации данных, однако в отличие от аналогичных систем авторы предлагают добавлять в обучающую выборку не все сгенерированные изображения, а только те, которые позволяют повысить точность повторной идентификации. Для этого отбрасываются изображения, которые имеют схожие признаки с ранее полученными, так как они могут снижать качество обучения, увеличивать время и при этом при-

водить к разбалансировке при обобщении. В этом случае система будет считать, что признаки, выделенные для схожих изображений, имеют большее значение, чем те, примеров которых было недостаточно. Для решения этой проблемы используется метод Local Outlier Factor (LOF), который контролирует количество схожих сгенерированных изображений и в случае, если их число возрастает, часть из них случайным образом отбрасывает. Такой подход позволяет не только повысить точность повторной идентификации, но и значительно улучшить устойчивость системы к смещению домена. В [55] приводятся результаты сравнения с другими алгоритмами, направленными на решение проблемы смещения домена, и в рейтинге точности в метриках Rank1, Rank5 и mAP предложенный в [55] подход занимает первые и вторые позиции для разных наборов данных среди современных подходов.

В [56] рассматривается подход, направленный на генерацию дополнительных изображений людей, когда в системе видеонаблюдения количество изображений с одной камеры больше, чем с другой, или вид с другой камеры для определенного человека отсутствует. Этот подход применяется для повышения робастности алгоритмов, если необходимы пары изображений одного и того же человека с разных камер. Однако такие образцы генерируются не в виде изображений, а в пространстве признаков. Это обусловлено тем, что при генерации изображений требуются значительно большие вычислительные затраты генеративной модели на качественное формирование фона и освещенности. Однако это не всегда оказывает положительное влияние на модель повторной идентификации, тогда как генерация только признаков не учитывает особенностей всего изображения снимаемой сцены.

4. Анализ используемых признаков

Для повторной идентификации с помощью СНС используются: глобальные признаки (рис. 6,*а*), т.е. формируемые для всего изображения человека в целом; локальные, когда изображение разделяется на отдельные фрагменты (рис. 6,*б*); ключевые точки (рис. 6,*в*), предполагающие для каждого участка изображения отдельный вектор признаков; дополнительные признаки (рис. 6,*г*), к которым можно отнести вспомогательные аннотации, информацию о времени и месте съемки, атрибуты; признаки человека из последовательности кадров (рис. 6,*д*).

4.1. Глобальные признаки

При повторной идентификации людей использование глобальных признаков является базовым подходом, и они применяются совместно с локальными [3] или дополнительными [57] для повышения точности повторной идентификации или в алгоритмах, в которых увеличение эффективности реидентификации достигается за счет их получения [58] или обработки [52].

При использовании глобальных признаков система повторной идентификации может оказаться недостаточно устойчивой к окклюзиям из-за того,

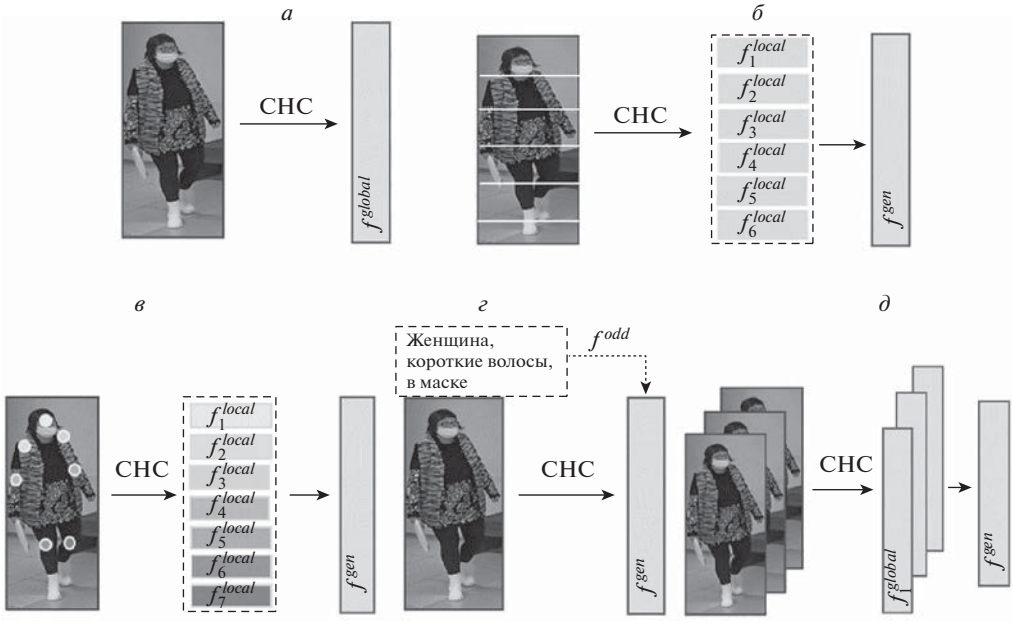

Рис. 6. Стратегии изучения и использования признаков.

что в сформированном векторе признаков для скрытого изображения часть дескрипторов будет характеризовать не внешность человека, а предмет, его перекрывающий. Кроме того, при таком подходе могут «теряться» признаки мелких отличительных деталей внешности, например таких, как очки, фурнитура одежды или сумки, которые могли бы служить характерным отличием при принятии решения о принадлежности обнаруженного человека к запросу.

4.2. Локальные признаки

Для снижения влияния недостатков глобальных признаков применяются локальные, которые могут рассматриваться как самостоятельно, так и в совокупности с глобальными. Например, в [59] предлагается горизонтальное разделение изображения на шесть равных частей и изучение каждой части в отдельности. Такой подход получил название Part-based Convolutional Baseline (PCB) и является надстройкой над CHC, при этом осуществляется разделение на части выходных данных первого сверточного слоя. Он позволяет повысить точность повторной идентификации на 1–2% в метриках Rank1 и mAP. Недостатком является требование к расположению и содержимому каждой части: человек должен находиться в строго вертикальном положении и фрагменты изображения должны располагаться в «правильных» местах. Ошибки обнаружения, когда часть человека оказывается обрезана ограничительной рамкой, могут приводить к ошибкам идентификации.

В [60] проводилось исследование по оценке влияния количества фрагментов, на которые разделяется изображение, на точность повторной идентификации. Изображение разбивалось на два, три, четыре, шесть, восемь и двенадцать фрагментов, и лучший результат точности повторной идентификации в метриках Rank1 и mAP был получен при делении изображения на шесть частей.

В [61] представлен алгоритм для реидентификации, основанный на рассмотрении ключевых частей тела человека. Так, с помощью HR-Net [62] извлекаются ключевые точки, а затем исследуются признаки в окрестностях каждой из них. Данный подход направлен на уменьшение влияния окклюзий. Поэтому при сопоставлении векторов признаков не учитываются дескрипторы ключевых точек, которые оказались скрыты.

В [63] рассматривается алгоритм, требующий разделения изображения фигуры человека на 6 горизонтальных частей, при этом сеть пытается предсказать, есть ли на каждой из них видимая часть фигуры человека. При положительном решении сети с помощью оценщика поз AlphaPose [64] определяются ключевые точки человека и при предсказании, является ли обнаруженный человек искомым, признаки невидимых частей не учитываются. Это позволяет повысить точность повторной идентификации человека и увеличить устойчивость системы к окклюзиям.

4.3. Дополнительные признаки

Еще одним подходом увеличения точности повторной идентификации является использование дополнительной информации, которая предоставляется с набором данных в виде аннотаций. Использование такого подхода предлагается в [57], при этом с помощью CHC (DenseNet-121, ResNet-50 или PCB) извлекаются визуальные признаки объектов, а номер камеры и номер кадра содержатся в названиях самих файлов. После ранжирования таблицы визуальных признаков из нее удаляются дескрипторы изображений, которые нерелевантны по пространственно-временным характеристикам людей, т.е. для тех, которые физически не могли находиться в определенном месте или в определенный час.

В большинстве случаев в алгоритмах повторной идентификации неочевидны типы признаков, используемых при принятии решения о сходстве или различии запроса и изображений людей в галерее. В [65] проводится исследование и предлагается подход, позволяющий определить и визуализировать признаки, которые система рассматривала при принятии решения, какие именно из них были значимыми и какой вклад вносит каждый атрибут. Для этого разработан метод, получивший название AMD (Attributeguided Metric Distillation), который представляет собой интерпретатор, подключаемый к целевой модели для оценки вклада каждого атрибута и визуализации наиболее значимых деталей. Интерпретатор учится разделять расстояние между признаками различных людей на основе атрибутов, и вводится

функция потерь, которая позволяет сосредоточиться на характерных отличиях. Эксперименты авторов показывают, что предоставляется возможность не только визуализировать значимые признаки, но и дополнительно улучшить точность повторной идентификации в целевых моделях. Представлены в данной работе также результаты исследования, показывающие улучшение точности повторной идентификации при тестировании алгоритма на междоменных данных.

В [66] предлагается повышение устойчивости систем реидентификации к смещению домена. Как правило, для таких систем предполагается, что есть исходный домен (sourse domain), используемый для обучения, и целевой (target domain), на котором осуществляется тестирование. При этом считается, что они изолированы между собой. В [66] применяются промежуточные домены в качестве дополнительной информации, которые позволят уменьшить различие между исходной и целевой областями. На вход базовой СНС подаются изображения как целевого, так и исходного домена, на их основе формируются дескрипторы, которые затем объединяются с различными соотношениями смешивания для получения вектора признаков промежуточного домена. Для этого применяется технология, предложенная в [67]. При объединении дескрипторов изображений из разных доменов возникает такой побочный эффект, как смешение признаков изображений разных людей и генерации изображения нового человека. Это может привести к тому, что в процессе обучения сеть сосредоточится на человеке со смешанными дескрипторами, вместо того чтобы учитывать разнообразие стилей в разных доменах. Для компенсации этого явления применяется дополнительный модуль, использующий подход к переносу стилей AdaIN [68], который позволяет получить дескрипторы одного и того же человека с учетом особенностей целевого или исходного домена. Сгенерированные признаки промежуточных доменов используются для обучения СНС и уменьшают расстояние между извлеченными дескрипторами из исходного и целевого доменов.

В [69] для решения таких проблем при повторной идентификации, как изменение освещения, окклюзии, фоновые помехи и возможная смена внешнего вида, предлагается использование технологии Wi-Fi, что позволяет подсчитывать и определять локализацию людей. Процедура обнаружения человека использует вариации Wi-Fi сигналов, которые могут информировать о присутствии человека и их можно отслеживать с помощью информации о состоянии канала (channel state information (CSI)) точек доступа. Из Wi-Fi сигнала извлекаются значимые признаки, на основе которых формируется радиобиометрическая подпись, используемая для реидентификации человека.

В [70] в качестве дополнительной информации, позволяющей повысить точность повторной идентификации в невидимых доменах, предлагается использовать «обучение распределению меток» (Label distribution learning (LDL)). Для обучения СНС используется несколько наборов данных, а сам процесс направлен на поиск взаимосвязи между изображениями разных людей. Каждый человек рассматривается как отдельный класс, и поиск соот-

ветствий между различными классами из разных наборов данных позволяет извлекать признаки, инвариантные к домену. Особое внимание уделяется похожим людям из разных доменов, что позволяет сформировать дескриптор, характеризующий внешность человека, а не условия видеонаблюдения. Для уменьшения разрыва между данными из разных доменов метки (идентификаторы) изображений для обучения распределяются таким образом, чтобы больше внимания уделять не самому домену, к которому принадлежит класс, а междоменным связям.

В [71] в качестве дополнительных признаков используется информация о ракурсе человека и при повторной идентификации учитываются признаки, связанные с углом обзора. С помощью СНС определяется один из трех рассматриваемых ракурсов, таких как вид спереди, сбоку и сзади, что позволяет улучшить устойчивость системы к смене доменов.

4.4. Признаки, использующие временные особенности

Алгоритмы повторной идентификации по последовательности кадров (video-based) используют преимущества временной составляющей, которой обладает видеоряд, в отличие от анализа отдельных кадров [60]. В [3] предлагается алгоритм, объединяющий как глобальные, так и локальные признаки на изображении человека для повышения точности повторной идентификации на видео. На разных уровнях пирамиды, представленной на рис. 7, изображение разделяется вертикальными или горизонтальными линиями и для каждого фрагмента изображения извлекается вектор признаков. Общий вектор признаков для каждого i-го человека в [9] определяется как

$$f_i^{gen} = \left[f_i^{global}; f_{i,v}^{local-vertical}; f_{i,h}^{local-hotizontal}; f_{i,patch}^{local-patch} \right], \tag{6}$$

где $v, h, patch$ — количество частей, на которые разделяется изображение на каждом уровне пирамиды.

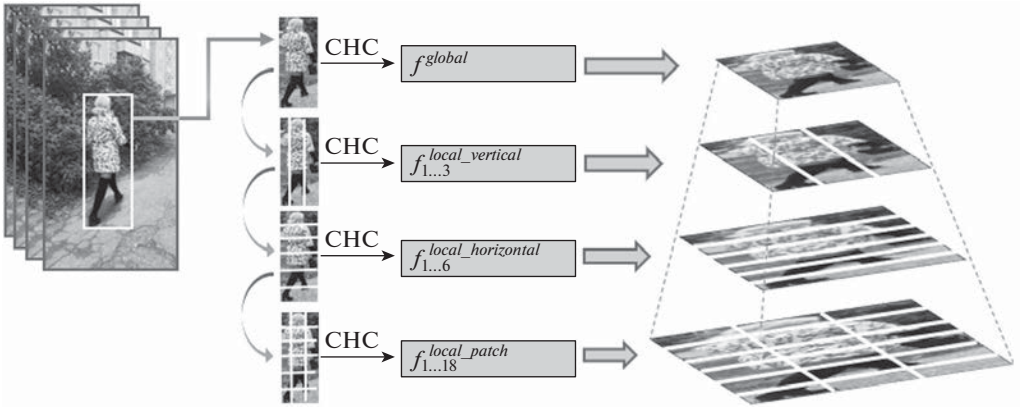

Рис. 7. Извлечение глобальных и локальных признаков на основе разделения изображения и многоуровневой пирамиды.

Для последовательности из K кадров видео вектор признаков для каждого человека описывается выражением

$$(7)\quad \overline{f}_i^{gen} = \left[\sum_{k=1}^{K} f_{i,k}^{global}; \sum_{k=1}^{K} f_{i,v,k}^{local-vertical}; \sum_{k=1}^{K} f_{i,h,k}^{local-hotizontal}; \sum_{k=1}^{K} f_{i,patch,k}^{local-patch} \right].$$

В [72] предлагается извлекать информацию о походке для силуэтов людей с использованием метода вычитания фона. Несмотря на то что цветные изображения содержат больше информации, чем образ фигуры человека, анализ силуэта позволяет сосредоточиться на определении особенностей, характерных для разных людей при движении. На первом этапе в [72] из видео удаляются фон и яркостно-цветовые отличия человека, в результате выделяется образ его фигуры. После вычитания фона генерируются ограничительные рамки для всех людей на каждом пятом кадре видео, а для расчета остальных ограничительных рамок используется линейная интерполяция. Извлеченные силуэты нормализуются по аналогии с методом, предложенным в [73], и на первом этапе рассматривается верхняя и нижняя часть фигуры, а затем анализируется совокупная сумма пикселей по оси X относительно центра этого объекта. После этого все изображения приводятся к единому размеру с сохранением соотношения сторон, но с высотой 224 пикселя. Согласно [72] для реидентификации по походке необходимо сформировать изображение (Gait Energy Images (GEI)), отражающее характерные особенности человека при ходьбе на основе анализа последовательности кадров. Для формирования GEI определяется траектория движения с использованием центральных координат ограничительных рамок. Полученная криволинейная траектория движения человека с помощью алгоритма кусочной регрессии разделяется на несколько прямолинейных участков. Для каждого такого участка к соответствующей последовательности кадров применяется алгоритм кластеризации k-средних, и формируется GEI.

В [74] рассматривается подход для повторной идентификации на видео, в котором к определенным последовательным кадрам применяется операция 3D свертки, объединяющая визуальную и временную составляющую, что позволяет учитывать изменения внешности в процессе движения. Кроме этого, используется специальная архитектура сети SSN для извлечения признаков отдельных частей тела и разделения дескрипторов на группы с учетом двигающихся и статических частей тела на видео.

В [75] предлагается выделять наиболее эффективные пространственно-временные признаки на основе анализа глобальных и локальных декрипторов для видеопоследовательности. Для построения глобальных признаков используется модуль Relation-Based Global Feature Learning Module (RGL), с помощью которого формируются карты корреляций дескрипторов между кадрами для поиска наиболее важных, а для синтеза локальных применяется модуль Relation-Based Partial Feature Learning Module (RPL), который позво-

ляет определить взаимосвязь между признаками одного и того же фрагмента на разных кадрах.

В [76] для более эффективного использования временной информации в видео предлагается подход, который включает два модуля. Первый Key Frame Screening with Index (KFSI) предполагает поиск похожих кадров и выбор из них для обучения СНС наиболее информативных для реидентификации. Второй модуль Feature Reorganization Based on Inter-Frame Relation (FRBIFR) предназначен для выявления наиболее значимых признаков людей на основе анализа их расположения на последовательности кадров, что позволяет уменьшить влияние шумовых факторов, например, перекрытий изображений людей.

4.5. Признаки ключевых областей

Для повышения устойчивости к влиянию помех фона и изменению признаков объекта при движении ряд исследователей предлагают выполнять поиск и выделение областей с использованием модулей (моделей, механизмов) внимания (attention module, attention model, attention mechanisms) [77]. В [78] для этого применяются локальный и глобальный анализы и предлагается модуль RGA (Relation aware global attention), который охватывает структурную информацию всего изображения и изучает фрагментарные отличительные особенности. Нахождение ключевых областей позволяет определить местоположение значимых отличительных признаков. Для их поиска выполняется попарное сравнение каждого дескриптора со всеми остальными и вычисленный результат включается в общий вектор признаков, позволяет учитывать взаимосвязть глобальных и локальных отличий изображений людей.

Механизм внимания используется во временной области [79], в [60, 80] анализируется пространственно-временная, может применяться в пространственно-локальной [81], и направлен на оценку позы человека и предсказание видимых частей. В [82] предлагается механизм самовнимания (self-attention) для повышения обобщающей возможности СНС путем учета взаимосвязи признаков.

Пирамидальный модуль для извлечения признаков с применением мультивнимания (pyramid multi-part features with multi-attention) (PMP-MA) рассмотрен в [60]. Полученные таким путем признаки позволяют учитывать важные отличительные особенности с различной степенью детализации. В [60] показана точность Rank5 = 99,3% на наборах данных iLIDS-vid и DukeMTMC-VideoReID, а для PRID Rank5 = 100%.

В [83] предлагается добавлять модули внимания между блоками ResNet для улучшения возможности извлечения признаков из кадров видеоряда. При прохождении изображения по СНС, часть важной информации может быть утеряна, но при этом сформированный вектор признаков будет содержать избыточную информацию для реидентификации. Поэтому в [83] предлагается встраивать модули пространственного внимания на разных уровнях сети

ResNet. Выходные карты признаков с определенных уровней СНС объединяются и формируют дескриптор для каждого отдельного кадра видеопоследовательности. Модуль внимания применяется для усреднения значений полученных карт признаков и построения результирующего вектора.

4.6. Метрики для определения расстояния между признаками

Для поиска изображения человека x_p в галерее $G = \{g_i | i = 1, \ldots, N\}$ из N изображений применяется вычисление расстояний между векторами признаков p-го запроса и изображения g_i. На данном этапе наиболее применимы следующие метрики:

1. Косинусное расстояние (Cosine distance) [57, 14]:

$$d(p, g_i) = \frac{x_p x_{g_i}}{\|x_p\| \|x_{g_i}\|}. \tag{8}$$

2. Расстояние Евклида (Euclidean distance) [7, 10, 13, 26, 84]:

$$d(p, g_i) = \|x_p - x_{g_i}\|_2^2. \tag{9}$$

3. Расстояние Махаланобиса (Mahalanobis distance) [85]:

$$d(p, g_i) = \sqrt{(x_p - x_{g_i})^T M^{-1} (x_p - x_{g_i})}, \tag{10}$$

где M — ковариационная матрица.

4. Расстояние Жаккара для k-ближайших соседей (Jaccard distance) [85]:

$$d(p, g_i) = 1 - \frac{|R^*(p, k) \cap R^*(g_i, k)|}{|R^*(p, k) \cup R^*(g_i, k)|}, \tag{11}$$

где $R^*(p, k)$ и $R^*(g_i, k)$ — множества ближайших соседей.

Следует отметить, что для повышения точности повторной идентификации в некоторых алгоритмах применяют повторное ранжирование после первой сортировки, которое позволяет уточнить результат. В [85] для первоначальной сортировки используется расстояние Махаланобиса. Из полученной таблицы выбираются первые k изображений и включаются в $R(p, k)$, а затем выполняется повторное ранжирование с использованием расстояния Жаккара.

В [26] на основе расстояния Евклида выполняется первичная сортировка векторов признаков. Далее при повторном ранжировании из полученной таблицы $S(p, g)$ выбираются k-первых результатов и для каждого из них осуществляется поиск в галерее. В результате формируются новые списки $S(r_i, g)$ с весовыми коэффициентами, которые определяются как $\frac{1}{i+1}$, где $i = 1, \ldots, k$. Итоговая таблица признаков вычисляется по формуле

$$S^*(p, g) = S(p, g) + \sum_{i=1}^{k} \frac{1}{i+1} S(r_i, g). \tag{12}$$

В [84] предлагается учитывать контекстную информацию ранжирования дескрипторов в процессе обучения СНС совместно с признаками для повторной идентификации. Алгоритм использует двухпоточную архитектуру, состоящую из внешнего и внутреннего потоков. На первом из них применяется сортировка для каждого запроса, что позволяет найти наиболее эффективные визуальные различия вверху ранжированного списка галереи и сформировать предварительный набор для дальнейшей обработки. На втором потоке анализируются локальные признаки для полученного результата предыдущего шага. Предполагается, что такой подход создает гибридное ранжирование для сопоставления людей, позволяющее повысить точность повторной идентификации по сравнению с методами, в которых применяется постобработка списка. Кроме указанных метрик, для оценки схожести признаков могут быть использованы и другие [86], однако эффективность их требует дополнительных исследований.

5. Модели и обучение СНС для описания изображений людей

5.1. Базовые СНС

Наиболее часто при реидентификации в настоящее время в качестве базовых СНС для извлечения признаков используются ResNet-50 [87] в работах [12, 65, 88] и DenseNet-121 [89] в работах [7, 57], а также MobileNetV2 [88, 90], PCB [57, 84], GoogleNet [91], или оригинальные архитектуры СНС, например, как в [92]. В [93] предлагается подход, который позволяет повысить устойчивость системы к окклюзиям. При этом повторная идентификация выполняется по изображению головы человека, а для обнаружения ограничительных рамок используется СНС YOLOv3.

Архитектуры семейства ResNet характеризуются наличием Res-блоков (рис. 8,*а*), которые используют пропуск соединений (scip-connection) для снижения вероятности возникновения исчезающих градиентов при обучении. Res-блок состоит из двух ветвей, одна из которых содержит сверточные слои, а другая передает информацию на выход без изменений. На выходе данные с обоих ветвей суммируются. В процессе обучения при обратном распространении ошибки такой подход не позволяет обнулить градиенты в СНС.

Архитектура DenseNet-121 (рис. 8,*б*) отличается наличием соединений между слоями, при которых карты признаков всех предыдущих слоев используются в качестве входных для всех последующих в блоке. Кроме этого, карты признаков не суммируются от слоя к слою, что характерно для ResNet, а конкатенируются. Некоторые исследователи приводят результаты сравнения работы предлагаемых алгоритмов с использованием в качестве базовых СНС для извлечения признаков различные типы архитектур. Так, в [7] выполняется сравнение эффективности ResNet-50 и DenseNet-121 и показано повышение точности в метриках Rank1 и mAP при использовании DenseNet-121. В [65] для реидентификации исследованы ResNet-34, ResNet-50 и ResNet-101

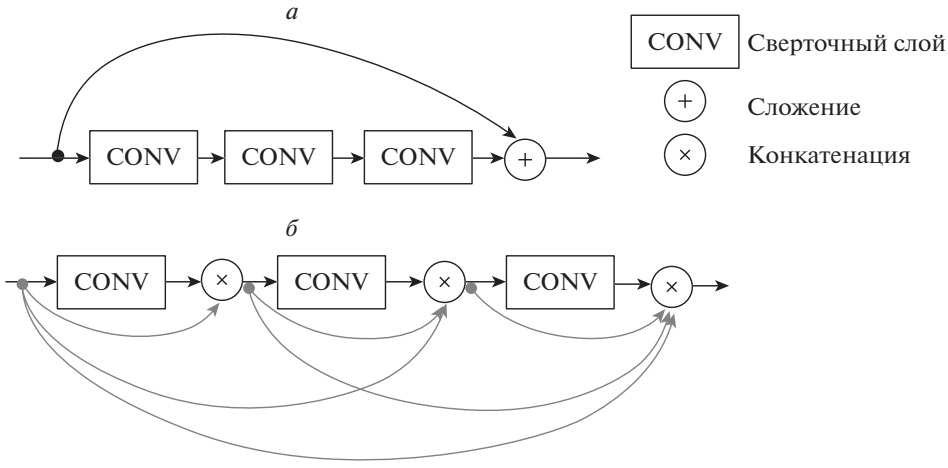

Рис. 8. Структуры блоков DenseNet и ResNet.

и представлено, что увеличение глубины сети положительно сказывается на точности повторной идентификации. В [57] выполнена оценка эффективности PCB [59], которая используется как надстройка для ResNet-50, ResNet-50 и DenseNet-121. Анализ результатов экспериментов показывает, что наилучшей в метриках Rank1 и mAP является PCB (Rank1 = 94,0, mAP = 82,8), более низкая точность для DenseNet-121 (Rank1 = 90,8, mAP = 76,9), а наименьшия значения метрик у ResNet-50 (Rank1 = 87,7, mAP = 72,2).

В [92] для реидентификации предлагается новая архитектура СНС SGWCNN (sparse graph wavelet convolution neural network) на основе анализа признаков последовательности кадров, что позволяет учитывать семантическую связь между локальными фрагментами людей на видео. Такой подход позволяет извлекать дополнительную информацию за счет пространственно-временного анализа видеоданных. Предполагается, что использование предложенной нейронной сети для уточнения региональных признаков позволяет более эффективно решать проблему кратковременных окклюзий при движении пешеходов.

Следует отметить, что качество работы СНС в значительной мере определяется гиперпараметрами при ее тренировке: количеством эпох, скоростью обучения, размером пакета изображений.

Количество эпох определяет, сколько раз каждое изображение из обучающей выборки пройдет по сети. При малых значениях данного параметра модель окажется не полностью обученной и в результате точность повторной идентификации будет низкой. Слишком большое количество эпох может привести к переобучению, т.е. сеть запомнит все рассмотренные изображения и не сможет эффективно обработать даже тестовые примеры. Для повторной идентификации тренировка СНС выполняется в большинстве случаев в течение 60–100 эпох. Как правило, на вход сети подаются пакеты с количеством изображений от 16 до 64. Увеличение размера пакета обусловлено стремле-

нием к распараллеливанию вычислений, так как это позволяет сократить время, затраченное на тренировку СНС, но снижает точность работы обученной нейронной сети. В [94] предлагается подход, согласно которому при тренировке СНС постепенно увеличивается размер пакета, что позволяет минимизировать уменьшение точности, обеспечивая сокращение времени обучения. Наиболее полное исследование влияния размера пакета на точность при тренировке СНС для реидентификации представлено в [60]. В данной работе показано, что наибольшей точности удалось достигнуть для пакета из 32 изображений на наборах данных DukeMTMC-VideoReID, MARS, iLIDS-vid, PRID.

Известно, что скорость обучения показывает, как изменяются весовые коэффициенты при каждом их обновлении. Для повторной идентификации при тренировке СНС используют планировщики скорости, которые позволяют изменять скорость обучения после некоторого интервала времени или по определенным критериям. В [95] рассматривается механизм снижения скорости ADEL, который отслеживает значения весов сети и каждый раз, когда они перестают изменяться скачкообразно, скорость обучения уменьшается. Это позволяет обеспечить более быструю сходимость в СНС.

В [96] предлагается подход, включающий три режима изменения скорости обучения η, которые зависят от кривизны λ_0 поверхности функции потерь. Первый режим предполагает медленную фазу (lazy phase), при ней скорость обучения имеет относительно небольшое значение $\eta < \frac{2}{\lambda_0}$ и шаг изменения весов остается практически постоянным на первом этапе обучения. Второй режим характеризуется быстрой фазой (catapult phase), при которой скорость обучения принимает значения $\frac{2}{\lambda_0} < \eta < \eta_{\max}$. На этом этапе наблюдается экспоненциальный рост потерь и быстрое уменьшение кривизны η до тех пор, пока не стабилизируется на значении $\lambda_{final} < \frac{\eta}{2}$. При соответствии этому условию достигается плоский минимум. Фаза расхождения (divergent phase) выполняется на третьем режиме. При этом скорость тренировки превышает значение $\eta_{\max}$ и модель перестает обучаться. Кроме этого, в [96] выдвигается предположение, которое затем подтверждается исследованиями, что использование больших скоростей обучения позволяет находить плоские минимумы, которые обобщают лучше, чем резкие. К этому же, по мнению авторов, приводит и использование небольших пакетов для обучения.

5.2. Модификации СНС

Изменения базовых архитектур предоставляют возможности для повышения точности работы систем повторной идентификации. В [88] исследуется влияние способа нормализации данных на выходе сверточных слоев и предлагается технология MetaBIN (Meta Batch-Instance Normalization), которая использует комбинацию двух подходов: пакетную нормализацию и нормализацию отдельных изображений [97]. Первый позволяет получать информацию о различных стилях изображений в пакете. Однако это может приводить к снижению точности реидентификации в невидимых доменах. Второй подход

позволяет игнорировать информацию об особенностях домена, однако недостатком является возможное уменьшение при этом полезной информации. Для решения двух этих проблем вводится обучаемый параметр, который позволяет найти баланс между рассмотренными подходами и тем самым не только повысить эффективность повторной идентификации, но и сделать систему более устойчивой при работе в другом домене. В [98] рассматривается влияние функции активации (ФА) в СНС ResNet-50, DenseNet-121 и DarkNet-53 на точность реидентификации. Наиболее распространенной функцией активации является ReLU [99], которая представляет собой кусочно-заданную функцию

$$\phi(x) = \begin{cases} x, & x > 0, \\ 0, & x \leqslant 0, \end{cases} \tag{13}$$

где x — входное значение нейрона.

Основное преимущество заключается в низкой вычислительной сложности как при прямом, так и при обратном проходе по сети. Однако значения производной на положительной части области определения функции активации могут приводить к взрывным градиентам при обучении, а на отрицательной — к потере некоторой информации при обучении, так как все нейроны с отрицательными значениями не будут активированы. Чтобы избежать этого, можно применять функцию Leaky-ReLU [100]

$$\phi(x) = \begin{cases} x, & x > 0, \\ \alpha x, & x \leqslant 0, \end{cases} \tag{14}$$

где α — угловой коэффициент, принимающий небольшие значения, традиционно $\alpha = 0{,}01$.

В [101] представлены результаты эмпирического исследования, в котором определяется влияние угла наклона отрицательной части функции на задаче классификации изображений при использовании ФА ReLU и Leaky-ReLU, а также их модификаций: параметрической выпрямленной линейной единицы (PReLU) и рандомизированной выпрямленной линейной единицы с утечкой (RReLU). Проведенные исследования показали, что лучшие результаты были получены при использовании PReLU. Однако в этом случае высока вероятность переобучения СНС при использовании небольшого набора данных, поэтому RReLU оказывается более эффективной на практике.

Кроме указанных модификаций, небольшой наклон в отрицательной части области определения функции имеют ФА ELU, SeLU, GeLU, что позволяет предположить эффективность их использования для повторной идентификации людей.

ФА ELU (Exponential Linear Unit) [102] определяется выражением

$$\phi(x) = \begin{cases} x, & x \geqslant 0, \\ \alpha(e^x - 1), & x < 0, \end{cases} \tag{15}$$

где $\alpha > 0$ — коэффициент, ограничивающий величину выходных значений на отрицательном участке области определения функции.

ФА SELU (Scaled Exponential Linear Unit) является масштабированным вариантом ELU и описывается выражением

$$\phi(x) = \lambda \begin{cases} x, & x \geqslant 0, \\ \alpha(e^x - 1), & x < 0. \end{cases} \tag{16}$$

В исследовании, представленном в [103], определяются значения для коэффициентов $\alpha = 1{,}67326$, $\lambda = 1{,}0507$.

ФА GELU (Gaussian Error Linear Units) [104] определяется выражением

$$\phi(x) = \frac{1}{2}x\left[1 + erf\left(\frac{x}{\sqrt{2}}\right)\right] \approx 0{,}5x\left(1 + \tan\left(\sqrt{\frac{2}{\pi}}(x + 0{,}044715x^3)\right)\right) \tag{17}$$

или

$$\phi(x) = x\sigma(1{,}702x), \tag{18}$$

где $\sigma = \frac{1}{1+e^{-x}}$ — функция активации сигмоиды.

В [105] для поиска наиболее эффективной ФА используется подход автоматической генерации, основанный на последовательном переборе унарных и бинарных функций, которые поочередно объединяются, а результат оценивается эмпирически. Полученая функция Swish определяется выражением

$$\phi(x) = x\sigma(\beta x), \tag{19}$$

где β — коэффициент, регулирующий степень кривизны функции, σ — функция сигмоиды.

В [106] рассмотрена ФА Mish

$$\phi(x) = x\tanh(softplus(x)) = x\tanh(ln(1 + e^x)). \tag{20}$$

ФА влияет как на динамику тренировки, так и на точность работы обученной модели. Из [98] следует, что использование вместо ReLU таких функций, как GeLU, Swish и Mish, может повысить точность повторной идентификации. Дополнительные исследования показали, что применение этих функций увеличивает время обучения модели, при этом не позволяет получить достаточно стабильный результат. К наиболее предпочтительным ФА для СНС при повторной идентификации можно отнести GeLU и ReLU.

Для решения специфических задач, например для неоднородных систем реидентификации [8], в которых используются изображения с инфракрасной камеры и с камеры видимого диапазона, предлагается новая архитектура СНС MCLNet (Modality Confusion Learning Network). MCLNet основывается

на частично разделенной двухпоточной сети. Для повышения устойчивости СНС к разнородным данным последовательно извлекаются признаки, характерные для каждого типа данных в отдельности, а затем общие дескрипторы. Так как видимые и инфракрасные образцы имеют разное распределение признаков и они не могут быть согласованы для сравнения, сеть обучается игнорировать информацию о модальности и пытается извлекать общие отличительные особенности для разнородных изображений человека. Чтобы не упустить важные особенности разных людей, создается механизм запутывания обучения, в результате чего несоответствие между разнородными изображениями сводится к минимуму, а сходство максимизируется. В [7] предлагается архитектура СНС RCSANet (Clothing Status Awareness Network) для долгосрочной повторной идентификации. Методы, применяемые для этого, учитывают, что после некоторого интервала времени человек сменил одежду, в которой он опять попадает в поле зрения видеокамеры. Однако такие подходы неэффективны, если в данном интервале времени человек не переоделся, и точность работы систем долгосрочной реидентификации значительно снижается. Для этого в [7] предлагается RCSANet, которая упорядочивает признаки пешеходов и включает в общий дескриптор особенности состояния одежды. RCSANet представляет собой двухпоточную систему, основанную на DenseNet-121, и содержит ICE-поток (Inter-Class Enforcement), который позволяет максимизировать различия для каждого человека, и ICR-поток (Intra-Class appearance Regularization), который используется для упорядочивания признаков, полученных в ICE, с учетом информации о том, имела ли место смена одежды. Предложенный подход для тестовой выборки, в которой смены одежды не было, позволил обеспечить значения Rank1 = 100% и mAP = 97,2%, а при наличии людей в различной одежде метрики равны Rank1 = 48,6% и mAP = 50,2%.

5.3. Сиамские сети

Сиамская нейронная сеть представляет собой такой тип архитектуры, который содержит две или больше идентичных подсетей с одинаковыми архитектурами, параметрами и весами. Выходом сиамской сети будет являться показатель подобия двух изображений, поданных на вход [107].

В сиамских сетях могут использоваться парные модели (рис. 9,*а*), состоящие из двух подсетей [108, 69], и триплетные [91], включающие три подсети (рис. 9,*б*).

В [108] сиамская архитектура используется для минимизации косинусного расстояния между признаками двух экземпляров при контрастном обучении для выявления сходства между ними. В [69] с помощью глубокой нейронной сети с двумя ветвями, работающей по сиамскому принципу, обрабатываются амплитуда и фаза Wi-Fi сигналов для извлечения значимых признаков радио-биометрической подписи, позволяющей повторно идентифицировать человека.

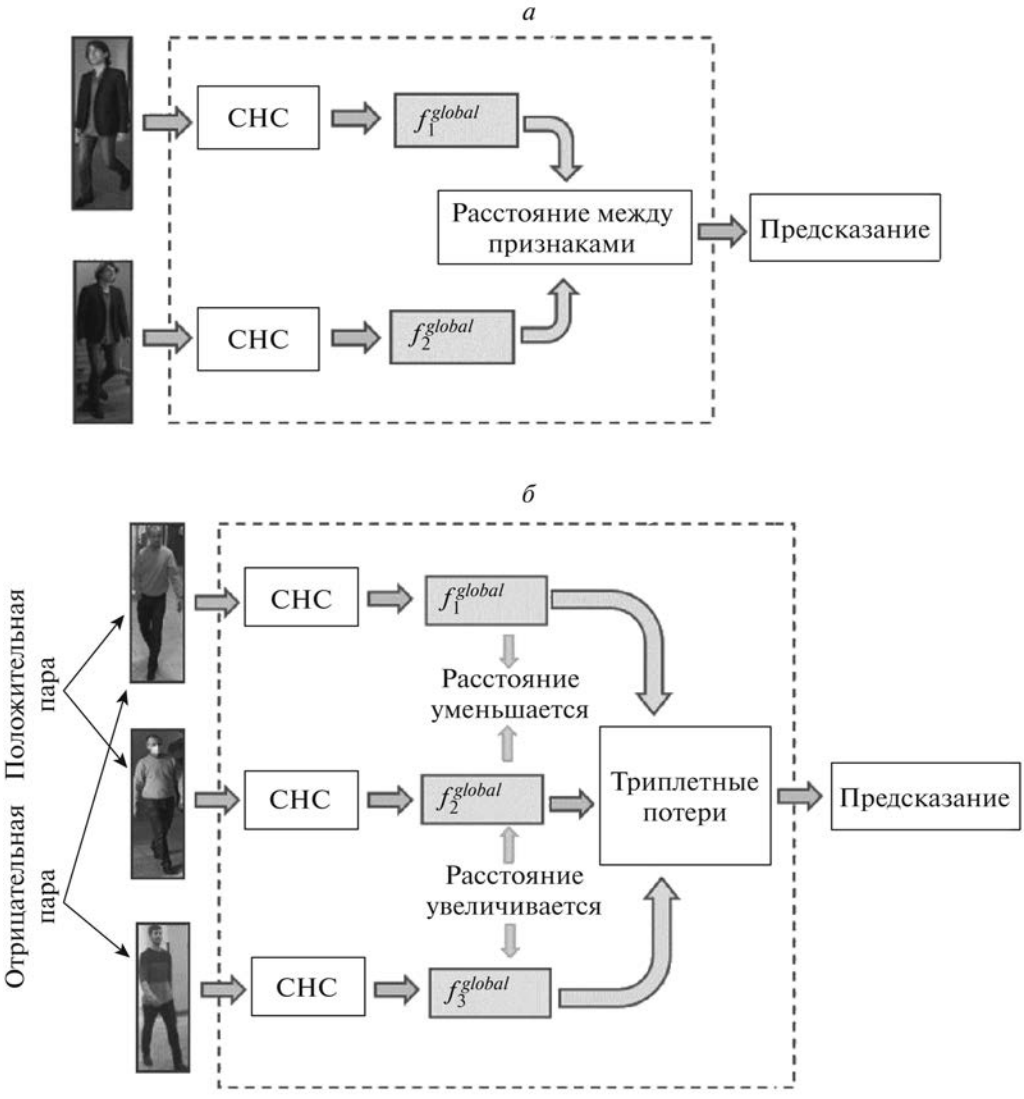

Рис. 9. Модели сиамской нейронной сети *а*) — парная модель; *б*) — триплетная модель.

В [109] сиамские сети используются для предотвращения переобучения и предлагается архитектура, состоящая из двух сиамских сетей. Первая из них является базовой, входными данными для нее служат положительные или отрицательные пары изображений людей. При этом положительной парой считаются изображения, полученные для одного человека в разное время, отрицательная пара представляет собой изображения двух разных людей. Признаки, извлеченные каждой ветвью базовой сиамской сети, подаются на входы другой сети, используемой для извлечения более глубоких признаков. Каждая из двух сиамских сетей предсказывает, является ли входная пара

изображениями одного человека, или нет. Между двумя сетями вводится функция потерь (verification loss), которая позволяет корректировать относительное расстояние между векторами признаков, полученными с каждой из сиамских сетей для людей с одинаковыми или разными идентификаторами, и тем самым улучшить точность идентификации.

В [110] предлагается глубокая архитектура для повторной идентификации, которая использует в структуре сиамской сети модуль внимания. Такой подход позволяет обеспечить согласованность важных деталей внешности человека с различных кадров и находить более важные отличительные черты для разных людей. Кроме этого, поиск расположения отличительных признаков на изображении реализуется в процессе обучения, что делает систему способной находить ключевые области автоматически.

Сиамская сеть с триплетными потерями предлагается в [91] с GoogleNet в качестве базовой подсети. Признаки людей при ее использовании извлекаются с разных уровней сети, а затем объединяются, формируя общую карту дескрипторов для каждого из входных изображений. Применение триплетных потерь позволяет приближать в пространстве признаков положительные пары изображений и отдалять отрицательные.

5.4. Обучение СНС

В общем обучение нейронной сети для эффективного извлечения признаков заключается в поиске весовых коэффициентов с целью уменьшения значения функции потерь L. Она отображает разницу между полученным результатом и ожидаемым. Для задачи повторной идентификации наиболее распространенными являются кросс-энтропийная функция потерь (Cross-entropy loss) [11, 14, 42, 57] и триплетные потери (Triplet loss) [7, 63, 111, 112].

Кросс-энтропийная функция потерь позволяет рассматривать реидентификацию как классификацию и используется после softmax-слоя [113]. Для набора из n тренировочных изображений $\{I_{ni}\}_{n_i=1,\ldots,n}$, содержащего n_{id} различных людей (классов) с соответствующими идентификационными метками $\{p_{n_i}^{ID}\}_{n_i=1,\ldots,n}$, $\{p_{n_i}^{ID}\} \in [1,\ldots,n_{id}]$, кросс-энтропийные потери можно рассчитать как [113]

$$L_i = -\sum_{k=1}^{n_{id}} \{p_{n_i}^{ID} = k\} \log \frac{e^{\hat{p}_{n_i}^{ID_i}}}{\sum\limits_{l=1}^{n_{id}} e^{\hat{p}_{n_i}^{ID_i}}}, \tag{21}$$

где $\hat{p_{n_i}^{ID_i}}$ — предсказанное значение.

Отличительной чертой триплетных потерь является рассмотрение двух пар изображений: при положительной паре изображения принадлежат одному и тому же человеку ($y_a = y_p$, где y_a — изображение человека с меткой идентификатора a, y_p — изображение, составляющее положительную пару, $p = a$); при отрицательной паре два изображения принадлежат разным людям ($y_a \neq y_n$, где y_n — изображение, составляющее отрицательную пару, т.е.

их идентификаторы не равны $n \neq a$). Таким образом, учитываются расстояние $d_{a,p}$ между признаками для положительной пары и расстояние $d_{a,n}$ между признаками разных людей. Чтобы СНС не только увеличивала $d_{a,n}$ для разных классов, но и уменьшала его для одинаковых, вводится коэффициент регуляризации m. Если не использовать этот коэффициент, то при обучении сеть будет увеличивать расстояние между изображениями разных людей и не учитывать расстояние между одинаковыми классами. Это связано с тем, что найти разницу между различными людьми легче, чем сходство между одинаковыми, соответственно m позволяет ограничить рост $d_{a,n}$ и обеспечивает уменьшение $d_{a,p}$. Для вычисления триплетной функции потерь используется выражение [111]

$$L = \sum_{\substack{a,p,n \\ y_a = y_p \neq y_n}} \max([m + d_{a,p} - d_{a,n}], 0). \tag{22}$$

В [53] при использовании триплетных потерь анализируется отрицательная пара, включающая изображения разных, но наиболее похожих людей. Такой подход позволяет научиться сети находить различия для людей со схожей внешностью. На наборе данных MSMT17 [19] метрика mAP составляет 84,4%, а в метрике Rank1 89,9%, что является лучшим результатом для MSMT17 на момент анализа.

В [114] предполагается, что при кластеризации изображений использование триплетных потерь является недостаточно эффективным подходом. Поэтому разработана функция потерь cluster loss, которая позволяет получить на выходе модели большие межклассовые и меньшие внутриклассовые различия, чем при использовании триплетов. Cluster loss вычисляется по формуле:

$$L_C = \frac{\beta \sum\limits_{i}^{p} d_i^{intra}}{\gamma + \sum\limits_{i}^{p} d_i^{inter}}, \tag{23}$$

где $d_i^{intra} = \sum_k \|f(x) - f_i^m\|_2^2$ — внутриклассовая вариация для каждого i-го идентификатора, представляющая собой расстояние между признаками $f(x)$ идентификатора из выборки и средним значением для этого идентификатора $f_i^m = \frac{\sum_K f(x)}{K}$ из K изображений; $d_i^{inter} = \sum_{\forall i_d \in P, i_d \neq i} \|f_i^m - f_{i_d}^m\|_2^2$ — межклассовая вариация, представляющая собой расстояние между средним значением признаков идентификатора и средним значением для признаков всех P идентификаторов.

Для повышения точности повторной идентификации иногда используют несколько функций потерь. Например, в [65] для определения наиболее эффективных признаков и наиболее значимых атрибутов предлагается две составляющие для функции потерь: метрического разделения (Loss function of metric distillation) L_d и приоритетных атрибутов (Loss function of attribute

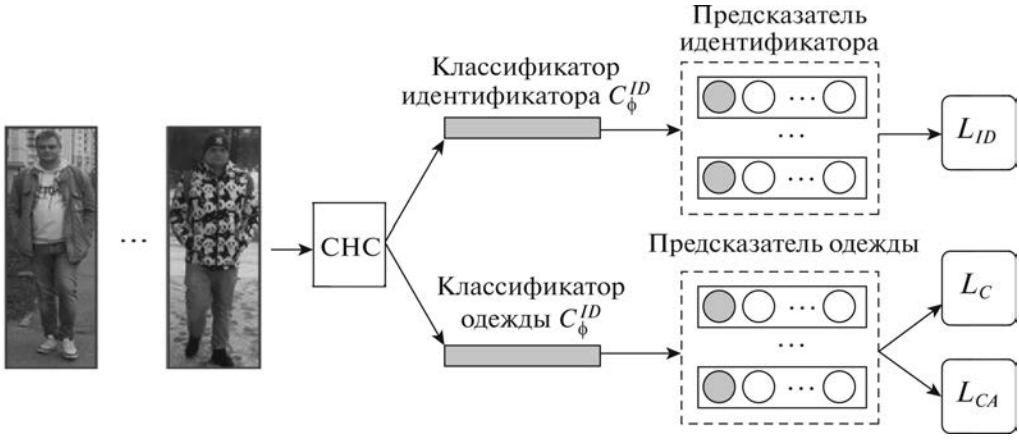

Рис. 10. Схема алгоритма Clothes-based Adversarial Loss.

prior) L_p:

$$L = L_d + \alpha L_{p1} + \beta L_{p2}, \tag{24}$$

где $L_d = |d_{i,j} - \sum_{k=1}^{M} d_{i,j}^k|$ — функция потерь, определяющая расстояние $d_{i,j}$ между векторами признаков, выделенных алгоритмом реидентификации для всего изображения, и признаками, выделенными модифицированным алгоритмом, направленным на поиск признаков различных атрибутов, что позволяет оценить вклад каждого атрибута в общий вектор признаков; d_{i,j^k} — расстояние между x_i и x_j для k-го из M атрибутов.

Очевидно, что составляющая функции потерь L_p состоит из двух частей: L_{p1} — определяет вклад общих атрибутов, L_{p2} — определяет вклад индивидуальных особенностей:

$$L_{p1} = \max\left(0, \left(\frac{M_E}{M}\right)^v - \sum_{e=1}^{M_E} \frac{d_{i,j}^e}{\hat{d}_{i,j}}\right) + \\ + \max\left(0, \sum_{c=1}^{M-M_E} \frac{d_{i,j}^c}{\hat{d}_{i,j}} - 1 + \left(\frac{M_E}{M}\right)^v\right), \tag{25}$$

$$L_{p2} = \sum_{e=1}^{M_E} \max\left(0, e^{-\lambda \frac{(\frac{M_E}{M})^v}{M_E}} - \frac{d_{i,j}^e}{\hat{d}_{i,j}}\right) + \\ + \sum_{c=1}^{M-M_E} \max\left(0, \frac{d_{i,j}^c}{\hat{d}_{i,j}} - e^{\lambda \frac{1-(\frac{M_E}{M})^v}{M-M_E}}\right), \tag{26}$$

где $\hat{d}_{i,j} \approx \sum_{k=1}^{M} d_{i,j}^k$ — предсказанное значение расстояния между признаками, M_E — количество уникальных атрибутов.

В [6] предлагается использовать состязательные потери на основе анализа одежды (Clothes-based Adversarial Loss (CAL)), которые позволяют извлекать признаки без учета одежды человека для долгосрочной повторной идентификации. Общая схема предлагаемого подхода показана на рис. 10 и состоит из двух классификаторов: идентификатора C_ϕ^{ID} и одежды C_ϕ^C. Каждый классификатор обучается отдельно. На первом этапе обучения минимизируются потери L_C (clothes classification loss) для классификатора одежды, которые основаны на кросс-энтропии между предсказанной меткой одежды $C_\phi^C(g_0(x_i))$ и y_i^C. При этом СНС тренируется не учитывать признаки одежды за счет минимизации функции потерь L_{CA} (Clothes-based Adversarial Loss), определяющей признаки, не относящиеся к одежде. После обучения классификатора одежды весовые коэффициенты для него фиксируются и дальнейший этап направлен на минимизацию функции потерь для классификатора идентификатора.

Тестирование предложенного алгоритма выполнялось на наборе данных CCVID [6], метод CAL позволил увеличить точность повторной идентификации более чем на 20% в метриках Rank1 и mAP по сравнению с базовым алгоритмом.

В настоящее время все большее внимание уделяется неконтролируемому или полу-контролируемому обучению, в которых данные не имеют заранее подготовленных меток и аннотаций. В некоторых работах по реидентификации людей исследователи предлагают использовать информацию, полученную с применением существующих размеченных наборов данных с известными идентификаторами, в невидимых доменах. Под невидимыми доменами подразумеваются наборы данных, изображения из которых не использовались при обучении и которые могут не иметь меток идентификаторов. В этом случае говорят о неконтролируемой адаптации домена (unsupervised domain adaptation (UDA)). Такой подход применяется в [115] и использует в качестве исходной информации объединение нескольких наборов данных. Как исходные и целевые домены рассматриваются Market-1501 [26], DukeMTMC-ReID [20], CUHK03 [18] и MSMT17 [19], которые объединяются в различных комбинациях. Кроме этого, предлагаются два модуля, позволяющие изучать отличительные признаки, характерные для одного домена, и для объединенных доменов. В первом случае предлагается модуль пакетной нормализации RDSBN, позволяющий снизить влияние признаков, специфических для домена, и улучшить различимость черт лица. Во втором — используется объединение информации о доменах на основе сети Graph Convolution Network (GCN), направленное на уменьшение расстояния между признаками разных доменов. GCN используется для построения графа, объединяющего все экземпляры в домене, т.е. создается узел, обобщающий характерные признаки для каждого человека внутри домена, что позволяет определить глобальные дескрипторы для домена в целом.

Неконтролируемые методы повторной идентификации, основанные на адаптации домена для целевой области, зачастую хорошо работают только

на одном домене, для которого были адаптированы. Решение этой проблемы рассматривается в [116] и предлагается проведение постоянной неконтролируемой адаптации к новым данным, т.е. непрерывное обучение. При этом знания, полученные ранее для предыдущих доменов, сохраняются. Это крайне важно в системах, применяемых в реальных условиях, когда новые данные появляются регулярно. Причем в систему могут быть включены дополнительные видеокамеры, установленные в других местах, а адаптация к новым данным должна быть с сохранением навыков повторной идентификации в уже известных доменах. Для решения этой задачи небольшое количество образцов из существующих доменов хранится в буферах долговременной памяти для сформированных ранее кластеров. В процессе адаптации модели к новой области старые образцы также добавляются в выборку, на основе которой выполняется контрастное обучение. Основная идея такой тренировки СНС заключается в максимизации сходства между положительной парой изображений, полученных при различных условиях.

При неконтролируемом обучении некоторые исследователи используют методы кластеризации для создания псевдометок во время обучения модели. При этом может возникать ситуация, когда в одном кластере объединяются изображения разных людей, а для одного и того же человека кластер может быть разбит на две группы. Это значительно снижает точность СНС, обученной на таких данных. В [117] предполагается, что из-за ограниченного количества выборок для каждого человека часть информации может отсутствовать. Поэтому предлагается метод Implicit Sample Extension (ISE), позволяющий создавать на границах кластеров образцы поддержки на основе реальных изображений текущего и соседних с ним кластеров через стратегию progressive linear interpolation (PLI), которая позволяет объединить два кластера, если они содержат изображения одного и того же человека, и разделить кластер, если в нем имеются изображения разных людей.

Методы самоконтролируемого обучения (self-supervised learning (SSL)) направлены на изучение отличительных признаков на основе больших массивов неразмеченных данных. В [118] для повышения точности повторной идентификации предлагается применение самоконтролируемого предварительного обучения с использованием немаркированных изображений людей, которое показывает лучшие результаты по сравнению с традиционным предварительным обучением на ImageNet. Основной идеей метода из [118] является выделение глобальных и локальных признаков. Структура предложенной в работе системы Part-Aware SelfSupervised pre-training (PASS) состоит из двух частей, имеющих одинаковую архитектуру: сеть учеников (student network) и сеть учителей (teacher network). PASS обучает сеть учеников соответствовать выходным данным учителя. На вход этой сети передаются как глобальные, так и локальные признаки, полученные для случайно выбранных областей. Сеть учителей анализирует только глобальные признаки. Сходство между результатами оценивается на основе кросс-энтропии. После предварительного обучения PASS имеет возможность изучать глобальные признаки и при

этом автоматически фокусироваться на различных локальных особенностях изображений.

В [119] рассматривается предварительное обучение для повторной идентификации, при котором для видео из набора LUperson применяется алгоритм сопровождения людей. Каждому сопровождаемому человеку присваиваются метки, которые используются для формирования новой обучающей выборки LUperson-NL и, соответственно, предварительной тренировки СНС. За счет такого подхода в LUperson-NL заносятся шумные метки (noisy labels), которые могут содержать ошибки, возникающие при присвоении идентификаторов в неразмеченных наборах данных. Примером является присвоение одному и тому же человеку, изображения которого получены с разных камер или в различное время, отличающихся идентификаторов. Другим примером является назначение одинаковых идентификаторов схожим по внешним признакам, но разным людям. Предполагается, что впоследствии ошибочные метки будут корректироваться. Подход из [119] предполагает три этапа. На первом выполняется контролируемое обучение повторной идентификации с использованием полученных шумных меток. Второй этап применяет контрастное обучение, которое позволяет исправить зашумленные метки. Для их исправления выбирается изображение-прототип, и по мере поступления новых данных, в случае их сходства с выбранным прототипом, изображения добавляются в кластер, и вычисляется усредненный для всех изображений в кластере вектор признаков, который динамически обновляется. На третьем этапе применяется контрастное обучение на основе уже исправленных меток. В результате похожие примеры объединяются в один прототип, а зашумленные метки исправляются.

В [120] предлагается подход PPLR (Part-based Pseudo Label Refinement), который уменьшает влияние шумных меток при неконтролируемом обучении за счет использования взаимодополняющей связи между глобальными и локальными признаками человека на изображении. Чтобы исключить влияние нерелевантных частей изображения, таких как окклюзии, которые могут появляться в разное время, искажать состав локальных и глобальных признаков человека и в итоге приводить к неверным предсказаниям, метки уточняются на основе введенного показателя взаимного согласования (cross agreement score) сходства k-ближайших соседей между пространствами глобальных и локальных признаков изображения человека.

В [108] рассмотрен подход для неконтролируемой повторной идентификации на основе скелета человека (skeleton-based) и предлагается схема контрастного обучения для извлечения признаков немаркированных 3D-скелетов человека. На последовательности необработанных данных накладываются маски скелетов, выбирается маска-прототип и по наиболее характерным особенностям скелета выполняется кластеризация. Чтобы найти отличительные особенности для различных прототипов без использования каких-либо меток, сопоставляются схожие черты скелетов. Чтобы учитывать корреляцию внутри одной и той же видеопоследовательности, связанную с изменениями

в процессе движения человека, используют сиамскую архитектуру, которая позволяет зафиксировать наиболее характерные признаки для каждого человека на основе его скелета.

6. Сравнение эффективности алгоритмов повторной идентификации на разных наборах данных

Сравнение эффективности различных алгоритмов повторной идентификации по одиночным изображениям на наиболее крупных и распространенных наборах данных Market-1501, DukeMTMC-ReID, MSMT17 представлено в табл. 2.

Таблица 2. Точность повторной идентификации по изображениям для наборов данных Market-1501, DukeMTMC-ReID и MSMT17

Алгоритм	Год	Метрики	Набор Market-1501	Набор DukeMTMC-ReID	Набор MSMT17
PCP+RPP [59]	2018	mAP	81,6	69,2	–
		Rank1	93,8	83,3	–
CASN [110]	2019	mAP	82,8	73,7	–
		Rank1	94,4	87,7	–
MGN+PTL [58]	2019	mAP	87,34	79,16	–
		Rank1	94,83	89,36	–
st-ReID [57]	2019	mAP	**95,5**	**92,7**	–
		Rank1	**98,0**	**94,5**	–
HOReID [61]	2020	mAP	84,9	75,6	–
		Rank1	94,2	86,9	–
AGW[2]	2021	mAP	–	–	49,3
		Rank1	–	–	68,3
		mINP	–	–	14,7
CIL [121]	2021	mAP	84,04	–	52,4
		Rank1	93,38	–	76,1
		mINP	57,9	–	12,45
SBS [65]	2021	mAP	88,29	78,26	–
		Rank1	95,55	89,21	–
Алгоритм из [55]	2021	mAP	89,2	79,6	57,2
		Rank1	96,2	91,0	81,9
FlipReID [52]	2021	mAP	94,7	90,7	81,3
		Rank1	95,8	93,0	87,5
HBReID [53]	2021	mAP	–	–	**84,4**
		Rank1	–	–	**89,9**
RANGEv2 [84]	2022	mAP	86,8	78,2	51,3
		Rank1	94,7	87,0	76,4
RANGEv2+K-reciprocal [84]	2022	mAP	91,3	84,2	–
		Rank1	95,1	88,7	–
CAL [6]	2022	mAP	87,5	–	57,3
		Rank1	94,7	–	79,9

В незначительном количестве работ [58, 59, 110] авторами приводятся результаты экспериментов для CUHK03. Однако данный набор включает не более пяти изображений для человека, полученных с каждой камеры из двух используемых, что является недостаточным для эффективной оценки точности реидентификации. Из табл. 2 очевидно, что при применении одного и того же алгоритма для разных наборов данных получены разные результаты точности: для набора Market-1501 наиболее высокие показатели, а для MSMT17 наиболее низкие значения. Такие результаты связаны с тем, что MSMT17 имеет значительно большее количество изображений (см. табл. 1), чем другие, и при его формировании использован более сложный сценарий видеонаблюдения, который охватил различное время суток и погодных условий, изображения как с внутренних, так и наружных камер видеонаблюдения. Хотя показатели точности для этого набора данных значительно ниже, чем для других рассматриваемых, но их можно считать более объективными, так как сценарий формирования MSMT17 более приближен к реальным ситуациям при реидентификации, следовательно, корректнее отображает эффективность алгоритмов для открытых систем повторной идентификации. На момент анализа лучший результат в метрике mAP при тестировании на наборе данных MSMT17 получен для алгоритма HBReID [53], для него Rank1 = 89,9%.

Это обеспечено за счет тщательного отбора положительных и отрицательных пар изображений для триплетных потерь и использования состязательных потерь, позволяющих изучить фон и не учитывать его при классификации людей, используя только признаки человека. Наиболее близкий по точности результат, Rank1 = 87,5, был получен с применением алгоритма FlipReID [52]. Отличием данного алгоритма является использование усредненных признаков входного и повернутого случайным образом изображения.

Для наборов Market-1501 и DukeMTMC-ReID среди рассмотренных алгоритмов повторной идентификации наиболее эффективный предложен в [57].

В данном алгоритме, кроме визуальных признаков, использовалась дополнительная информация о времени (номер кадра) и месте (идентификатор камеры) съемки, которая предоставляется с этими наборами данных в виде имени файла изображения из Market-1501 и DukeMTMC-ReID. При реидентификации авторы использовали условие о том, что человек не может находиться в поле зрения нескольких непересекающихся камер одновременно, ему требуется время для перехода. Таким образом, все изображения, имеющие визуальное сходство с запросом, должны быть проверены, могли ли эти люди быть в том или ином месте в определенное время относительно предыдущего, и все нерелевантные изображения по данным критериям не учитываются при реидентификации. Несмотря на то что данный подход был предложен в 2019 г., в настоящее время в метриках Rank1 и Rank5 этот подход имеет одни из лучших значений при тестировании на Market-1501 и DukeMTMC-ReID. Следует отметить, что авторы алгоритма из [57] не проводили эксперименты на других наборах данных, вероятно, из-за того, что с другими базами

Таблица 3. Точность алгоритмов повторной идентификации по видеопоследовательностям для наборов данных MARS, DukeVideo, PRID, QMUL iLIDS, iLIDS-VID, VIPer

Алгоритм	Год	Метрики	Набор MARS	Набор DukeVideo	Набор PRID	Набор iLIDS-VID
AGW [2]	2021	mAP	83,0	94,9	–	–
		Rank1	87,6	95,4	94,4	–
		mINP	63,9	91,9	95,4	–
PiT [3]	2022	mAP	**97,23**	–	–	86,80
		Rank1	**90,22**	–	–	92,07
MetaBin [88]	2021	mAP	–	–	81,0	87,0
		Rank1	–	–	74,2	81,3
SSN3D [74]	2021	mAP	86,2	**96,3**	–	–
		Rank1	90,1	96,8	–	88,9
PMP–MA [60]	2022	mAP	88,1	**96,3**	99,3	95,3
		Rank1	90,6	**97,2**	**98,9**	92,8
		Rank5	99,6	99,3	**100**	99,3
Алгоритм из [83]	2022	mAP	82,6	94,2	–	–
		Rank1	88,2	95,4	96,6	89,3
		Rank5	96,5	99,3	**100**	98,7

изображений не предоставлялось достаточно пространственно-временной информации. Анализ табл. 2 показывает, что несмотря на отличие показателей точности для различных наборов данных, в большинстве современных алгоритмов, если улучшение точности отмечается для одного из наборов данных, то вероятно, что и на других наборах данных при тетировании будет отмечаться увеличение точности.

Сравнение эффективности алгоритмов повторной идентификации по видеопоследовательностям представлено в табл. 3 и показывает, что для таких систем улучшение точности на одном наборе данных не всегда приводит к улучшению при тестировании на другом.

Наборы данных, используемые при обучении и тестировании, содержат последовательности изображений людей из нескольких кадров (треклеты), причем количество треклетов в отдельных наборах данных для каждого человека отличается количеством изображений. Применение треклетов позволяет учитывать временные признаки, что дает возможность исключить влияние кратковременных окклюзий, учесть информацию о походке или усреднить визуальные признаки для меняющих положение в пространстве частей тела на нескольких кадрах. Следовательно, количество изображений человека в треклете оказывает достаточно существенное влияние на оценку алгоритма.

При тестировании на наборе данных MARS (см. табл. 3) наилучшей точностью характеризуется алгоритм PiT [3] в метриках mAP и Rank1. Данный алгоритм использует пирамиды локальных признаков с разной степенью детализации и усреднение дескрипторов по нескольким кадрам. При тестиро-

Таблица 4. Точность повторной идентификации для алгоритмов междоменной повторной идентификации для наборов данных Market-1501, DukeMTMC-ReID, MSMT17, VIPeR, PRID и GRID

Алгоритм (Год)	Набор данных	Метрика	Тестовая выборка					
			Market -1501	DukeMT MC-ReID	MSMT17	VIPeR	PRID	GRID
Алгоритм из [48] (2020)	RandPerson [48]	mAP	28,8	27,1	6,3	–	–	–
		Rank1	55,6	47,6	20,1	–	–	–
	RandPerson [48] + MSMT17	mAP	35,8	39,8	36,8	–	–	–
		Rank1	62,3	61,0	65,0	–	–	–
SNR [12] (2020)	Market–1501	mAP	84,7	33,6	–	42,3	42,2	36,7
		Rank1	**94,4**	55,1	–	32,3	30,0	29,0
	DukeMTMC–ReID	mAP	33,9	72,9	–	41,2	45,4	35,3
		Rank1	66,7	84,4	–	32,6	35,0	26,0
	Market–1501+DukeMTMC–ReID+CUHK+MSMT17	mAP	82,3	73,2	–	65,0	60,0	41,3
		Rank1	93,4	85,5	–	55,1	49,0	30,4
NRMT[10] (2020)	Market–1501	mAP	–	62,3	19,8	–	–	–
		Rank1	–	78,1	43,7	–	–	–
	DukeMTMC–ReID	mAP	72,2	–	20,6	–	–	–
		Rank1	88,0	–	45,2	–	–	–
Алгоритм из [11] (2020)	Market–1501	* mAP	–	65,2	20,4	–	–	–
		Rank1	–	79,5	43,7	–	–	–
	DukeMTMC–ReID	mAP	71,5	–	24,3	–	–	–
		Rank1	88,1	–	51,7	–	–	–
CBN [9] (2021)	UnrealPerson [9]	mAP	54,3	49,4	15,3	–	–	–
		Rank1	79,0	69,7	38,5	–	–	–
JVTC [9] (2021)		mAP	80,2	**75,2**	34,8	–	–	–
		Rank1	93,0	**88,3**	68,2	–	–	–
MetaBin [88] MobileNetV2 (2021)	CUHK02+CUHK03+Market–1501 +DukeMTMC–ReID+CUHK–SYSU	mAP	–	–	–	66,0	79,8	**58,1**
		Rank1	–	–	–	56,9	72,5	**49,7**
MetaBin [88] ResNet–50 (2021)		mAP	–	–	–	**68,6**	**81,0**	57,9
		Rank1	–	–	–	**59,9**	**74,2**	48,4
QAConv [49] (2022)	ClonedPerson [49]	mAP	21,8	–	18,5	–	–	–
		Rank1	22,6	–	49,1	–	–	–
IDM [66] (2022)	Market–1501	mAP	–	73,2	40,2	–	–	–
		Rank1	–	85,5	**69,9**	–	–	–
	DukeMTMC–ReID	mAP	**85,3**	–	**40,5**	–	–	–
		Rank1	94,2	–	69,5	–	–	–
	MSMT17	mAP	85,2	73,6	–	–	–	–
		Rank1	94,1	84,6	–	–	–	–

вании на наборах данных Duke-Video и iLIDS-VID наилучшие результаты получены для алгоритма PMP-MA [60] в метрике mAP. В основе PMP-MA используется пирамидальное представление схем мультивнимания и учитываются результаты точной настройки СНС, в том числе подбор размера пакета данных при обучении. PMP-MA и алгоритм, предложенный в [83] при тестировании на базе данных PRID, позволяют получить Rank5 = 100%, что обеспечивается относительной легкостью набора для данной метрики, так как использовалось только две камеры, фон довольно равномерный, а окклюзии для человека с другими людьми встречаются редко.

Актуальным является сравнение точности работы алгоритмов реидентификации при обучении и тестировании на разных базах данных, что позволяет оценить эффективность при смене доменов. В табл. 4 представлена точность повторной идентификации алгоритмов, в которых для доменной переносимости использовались подходы, направленные на поиск признаков с учетом данной задачи.

Следует отметить, что независимо от используемого алгоритма увеличение обучающей выборки путем объединения существующих наборов данных позволяет повысить точность повторной идентификации, что подтверждается исследованиями в [12, 48]. В [48] добавление к синтетическому набору, используемому в качестве обучающей выборки, изображений из MSMT17 позволило повысить mAP с 47,6% до 61% при тестировании на DukeMTMC-ReID.

Включение в обучающую выборку данных из целевого домена позволяет увеличить Rank1 для MSMT17 с 6,3 до 36,8% в [48]. Аналогично, в [12] использование при обучении изображений из целевого домена позволяет увеличить mAP для Market-1501 с 33,9 до 82,3%, для DukeMTMC-ReID также обеспечивается значительное увеличение mAP с 33,6 до 73,2%.

Среди современных алгоритмов для междоменной повторной идентификации (см. табл. 4) в метрике mAP наибольшая точность достигнута для алгоритма IDM [66] при тестировании на наборах данных Market-1501 и MSMT17. В данном алгоритме для повышения устойчивости к смещению домена применялась генерация промежуточных доменов, которые объединяли в себе особенности исходного и целевого. При использовании в качестве целевого домена DukeMTMC-ReID наиболее эффективным оказался подход JVTC [9] с применением в качестве обучающей выборки синтетического набора данных UnrealPerson [9].

Алгоритмы реидентификации по видеопоследовательностям, направленные на повышение устойчивости к смещению домена, наиболее часто используют наборы данных VIPeR, PRID и GRID. Среди проанализированных подходов лучшие показатели Rank1 и mAP были получены при применении алгоритма MetaBIN [88] (см. табл. 4), основная идея которого заключается в обобщении слоев нормализации и снижении влияния особенностей, присущих исходному домену.

Таблица 5. Точность алгоритмов повторной идентификации с неконтролируемым и полуконтролируемым обучением для наборов данных Market-1501, DukeMTMC-ReID и MSMT17

Алгоритм	Год	Метрики	Набор Market-1501	Набор DukeMTMC -ReID	Набор MSMT17
[115] обучен на Market–1501	2021	mAP Rank1	– –	66,6 80,3	34,9 64,7
[115] обучен на Duke–MTMC–ReID	2021	mAP Rank1	81,5 92,9	– –	33,6 64,0
[116] с непрерывной адаптацией домена	2022	mAP Rank1	59,3 82,7	– –	40,8 67,5
ISE[117] + GeM–pooling	2022	mAP Rank1	85,3 94,3	– –	37,0 67,6
PASS [118]	2022	mAP Rank1	**93,3** **96,9**	– –	**74,4** **89,7**
PNL[119]+MGN пред–обучен на LuPerson	2022	mAP Rank1	91,9 96,6	**84,3** **92,0**	68,0 86,0
PPLR [120] без использования меток камеры	2022	mAP Rank1	81,5 92,8	– –	31,4 61,1
PPLR [120] с использованием меток камеры	2022	mAP Rank1	84,4 94,3	– –	42,2 73,3

Другим подходом для адаптации к смене домена при реидентификации является неконтролируемое или полуконтролируемое обучение на неразмеченных данных. Как очевидно из табл. 5, среди рассмотренных алгоритмов наилучшие предложены в [118, 119]. В [118] используются предварительное обучение на немаркированных изображениях людей, двухпоточная архитектура, глобальные и локальные признаки. Алгоритм из [119] предполагает предварительное обучение на немаркированном наборе данных LUPerson, для которого формируются и исправляются в процессе обучения шумные метки. В большинстве работ для предварительного обучения применяется набор данных ImageNet, однако последние исследования [118, 119] показали, что наиболее эффективно на этом этапе использовать изображения людей.

7. Заключение

Повторная идентификация человека в распределенной системе видеонаблюдения является достаточно новой актуальной задачей, которая в последнее время стала успешно решаться с помощью технологий глубокого обучения. В работе рассмотрены общие принципы организации повторной идентификации людей с использованием сверточных нейронных сетей при видеонаблюдении. Предложена классификация систем реидентификации. Приведен

анализ существующих наборов данных для обучения глубоких нейронных архитектур, описаны подходы для увеличения количества изображений в базах данных, рассмотрены виды признаков изображений людей. Представлен анализ основных применяемых для реидентификации моделей архитектур сверточных нейронных сетей, их модификаций, а также методов обучения. Проанализирована эффективность повторной идентификации людей на разных наборах данных, в том числе при междоменной реидентификации. Приведены результаты исследований по оценке эффективности повторной идентификации в различных метриках для существующих подходов, отмечены достоинства и недостатки разных метрик. Несмотря на то что глубокое обучение и нейронные сети продемонстрировали свои большие преимущества в анализе видеоизображений, все еще остаются проблемы, которые предстоит решить для повторной идентификации. Одним из наиболее заметных недостатков глубокого обучения является то, что для процесса обучения требуется огромное количество точных аннотированных наборов данных, что требует утомительной работы и часто приводит к искажениям. Многие исследователи начали делиться своими данными на общедоступных платформах, что полезно для разработки единого оценочного индекса, однако некоторые наборы данных для реидентификации исключены из общего доступа, например DukeMTMC-ReID [20], а MTMC17 [19] не доступен в публичном доступе и может быть получен лишь после подписания соглашения с авторами об использовании только в исследовательских целях без передачи третьим лицам [122].

Глубокое обучение достигло удовлетворительных результатов в задачах классификации и сегментации изображений, однако для задачи повторной идентификации, особенно по последовательностям изображений, производительность глубокого обучения все еще недостаточна высока. Поэтому актуальным направлением является разработка новых решений с использованием СНС, обеспечивающих более высокую точность и скорость работы, особенно для междоменной повторной идентификации.

СПИСОК ЛИТЕРАТУРЫ

1. *Ye S., Bohush R.P., Chen H.* Person Tracking and Re-identification for Multicamera Indoor Video Surveillance Systems // Pattern Recognit. Image Anal. 2020. No. 30. P. 827–837. https://doi.org/10.1134/S1054661820040136

2. *Ye M., Shen J., Lin G., Xiang T., Shao L., Hoi S.C.* Deep Learning for Person Re-identification: A Survey and Outlook // IEEE Transactions On Pattern Analysis And Machine Intelligence. 2021. https://doi.org/10.1109/TPAMI.2021.3054775

3. *Zang X., Li G., Gao W.* Multi-direction and Multi-scale Pyramid in Transformer for Video-based Pedestrian Retrieval // ArXiv, abs/2202.06014. 2022. https://doi.org/10.1109/TII.2022.3151766

4. *Mihaescu R., Chindea M., Paleologu C., Carata S., Ghenescu M.* Person Re-Identification across Data Distributions Based on General Purpose DNN Object Detector // Algorithms. 2020. No. 13. 343. https://doi.org/10.3390/a13120343

5. *Liu H., Qin L., Cheng Z., Huang Q.* Set-based classification for person re-identification utilizing mutual-information // 2013 IEEE International Conference on Image Processing. 2013. P. 3078–3082. https://doi.org/10.1109/ICIP.2013.6738634.

6. *Gu X., Chang H., Ma B., Bai S., Shan S., Chen X.* Clothes-Changing Person Re-identification with RGB Modality // ArXiv, abs/2204.06890, 2022. https://doi.org/10.48550/arXiv.2204.06890

7. *Huang Y., Wu Q., Zhong Y., Zhang Z.* Clothing Status Awareness for Long-Term Person Re-Idenification // 2021 IEEE/CVF International Conference on Computer Vision, 2021. P. 11895–11904. https://doi.org/10.1109/ICCV48922.2021.01168

8. *Hao X., Zhao S., Ye M., Shen J.* Cross-Modality Person Re-Identification via Modality Confusion and Center Aggregation // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. P. 16383–16392. https://doi.org/10.1109/ICCV48922.2021.0160

9. *Zhang T., Xie L., Wei L., Zhuang Z., Zhang Y., Li, B. Tian, Q.* UnrealPerson: An Adaptive Pipeline towards Costless Person Re-identification // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. P. 11501–11510. https://doi.org/10.1109/CVPR46437.2021.01134

10. *Zhao F., Liao S., Xie G., Zhao J., Zhang K., Shao L.* Unsupervised Domain Adaptation with Noise Resistible Mutual-Training for Person Re-identification // ECCV 2020. Lecture Notes in Computer Science, 2020. V. 12356. P. 526–544. Springer, Cham. https://doi.org/10.1007/978-3-030-58621-8_31

11. *Luo C., Song C., Zhang Z.* Generalizing Person Re-Identification by Camera-Aware Invariance Learning and Cross-Domain Mixup // ECCV 2020. Lecture Notes in Computer Science, 2020. V. 12356. P. 224–241. Springer, Cham. https://doi.org/10.1007/978-3-030-58555-6_14

12. *Jin X., Lan C., Zeng W., Chen Z., Zhang L.* Style Normalization and Restitution for Generalizable Person Re-Identification // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. P. 3140–3149. https://doi.org/10.1109/cvpr42600.2020.00321

13. *Song J., Yang Y., Song Y., Xiang T., Hospedales T.M.* Generalizable Person Re-Identification by Domain-Invariant Mapping Network // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019. P. 719–728. https://doi.org/10.1109/CVPR.2019.00081

14. *Ihnatsyeva S., Bohush R., Ablameyko S.* Joint Dataset for CNN-based Person Re-identification // Pattern Recognition and Information Processing (PRIP'2021) Proceedings of the 15th International Conference, 21–24 Sept. 2021, Minsk, Belarus / United Institute of Informatics Problems of the National Academy of Sciences of Belarus. Minsk, 2021. P. 33–37.

15. *Liao S., Mo Z., Hu Y., Li S.* Open-set Person Re-identification // ArXiv, abs/1408.0872, 2014. https://doi.org/10.48550/arXiv.1408.0872

16. *Li W., Zhao R., Wang X.* Human Reidentification with Transferred Metric Learning // Proceedings of the 11th Asian conference on Computer Vision (ACCV). 2012. https://doi.org/10.1007/978-3-642-37331-2_3

17. *Li W., Wang X.* Locally Aligned Feature Transforms across Views // 2013 IEEE Conference on Computer Vision and Pattern Recognition, 2013. P. 3594–3601. https://doi.org/10.1109/CVPR.2013.461

18. *Li W., Zhao R., Xiao T., Wang X.* DeepReID: Deep Filter Pairing Neural Network for Person Re-identification // 2014 IEEE Conference on Computer Vision and Pattern Recognition, P. 152–159. https://doi.org/10.1109/CVPR.2014.27

19. *Wei L., Zhang S., Gao W., Tian Q.* Person Transfer GAN to Bridge Domain Gap for Person Re-identification // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018. P. 79–88. https://doi.org/10.1109/CVPR.2018.00016

20. *Ristani E., Solera F., Zou R.S., Cucchiara R., Tomasi C.* Performance Measures and a Data Set for Multi-target, Multi-camera Tracking // ArXiv, abs/1609.01775, 2016. https://doi.org/10.1007/978-3-319-48881-3_2

21. Exposing.ai. Duke MTMC. URL: https://exposing.ai/duke_mtmc

22. *Zheng L., Zhang H., Sun S., Chandraker M., Yang Y., Tian Q.* Person Re-identification in the Wild // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. P. 3346–3355. https://doi.org/10.1109/CVPR.2017.357

23. *Xiao T., Li S., Wang B., Lin L., Wang, X.* Joint Detection and Identification Feature Learning for Person Search // IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017. P. 3376–3385. https://doi.org/10.1109/CVPR.2017.360

24. *Zheng L., Bie Z., Sun Y., Wang J., Su C., Wang S., Tian Q.* MARS: A Video Benchmark for Large-Scale Person Re-Identification // ECCV 2016. Lecture Notes in Computer Science, V. 9910. P. 863–884. Springer, Cham. 2016. https://doi.org/10.1007/978-3-319-46466-4_52

25. *Song G., Leng B., Liu Y., Hetang C., Cai S.* Region-based Quality Estimation Network for Large-scale Person Re-identification // AAAI. ArXiv, abs/1711.08766. 2018. https://doi.org/10.48550/arXiv.1711.08766

26. *Zheng L., Shen L., Tian L., Wang S., Wang J., Tian, Q.* Scalable Person Re-identification: A Benchmark // IEEE International Conference on Computer Vision (ICCV), 2015. P. 1116–1124. https://doi.org/10.1109/ICCV.2015.133

27. *Gray D., Brennan S., Tao H.* Evaluating Appearance Models for Recognition, Reacquisition, and Tracking // IEEE Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance. 2007.

28. *Hirzer M., Beleznai C., Roth P.M., Bischof H.* Person Re-identification by Descriptive and Discriminative Classification // SCIA. Lecture Notes in Computer Science. 2011. V. 6688. P. 91–102, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21227-7_9

29. *Zheng W., Gong S., Xiang T.* UnrealPerson: An Adaptive Associating Groups of People // BMVC. 2009. https://doi.org/10.5244/C.23.23

30. *Karanam S., Gou M., Wu Z., Rates-Borras A., Camps O.I., Radke R.J.* A Systematic Evaluation and Benchmark for Person Re-Identification: Features, Metrics, and Datasets // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019. No. 41. P. 523–536. https://doi.org/10.1109/TPAMI.2018.2807450

31. *Ihnatsyeva S., Bohush R.* PolReID, 2021. URL: https://github.com/SvetlanaIgn/PolReID

32. *Li S., Xiao T., Li H., Zhou B., Yue D., Wang X.* Person Search with Natural Language Description // 2017 IEEE Conference on Computer Vision and Pattern

Recognition (CVPR). 2017. P. 5187–5196.
https://doi.org/10.1109/CVPR.2017.551

33. *Ding Z., Ding C., Shao Z., Tao, D.* Semantically Self-Aligned Network for Text-to-Image Part-aware Person Re-identification // ArXiv, abs/2107.12666, 2021

34. *Li X., Zheng W., Wang X., Xiang T., Gong S.* Multi-Scale Learning for Low-Resolution Person Re-Identification // 2015 IEEE International Conference on Computer Vision (ICCV). 2015. P. 3765–3773.
https://doi.org/10.1109/ICCV.2015.429

35. *Jing X., Zhu X., Wu F., Hu R., You X., Wang Y., Feng H. Yang J.* Super-Resolution Person Re-Identification With Semi-Coupled Low-Rank Discriminant Dictionary Learning // IEEE Transactions on Image Processing, 2015. No. 26. P. 1363–1378. https://doi.org/10.1109/TIP.2017.2651364

36. *Wu A., Zheng W., Yu H., Gong S., Lai J.* RGB-Infrared Cross-Modality Person Re-identification // IEEE International Conference on Computer Vision (ICCV). 2017. P. 5390–5399. https://doi.org/10.1109/ICCV.2017.575

37. *Nguyen T.D., Hong H.G., Kim K., Park K.R.* Person Recognition System Based on a Combination of Body Images from Visible Light and Thermal Cameras // Sensors (Basel, Switzerland). No. 17. 2017. https://doi.org/10.3390/s17030605

38. *Pang L., Wang Y., Song Y., Huang T., Tian, Y.* Cross-Domain Adversarial Feature Learning for Sketch Re-identification // Proceedings of the 26th ACM international conference on Multimedia. 2018. https://doi.org/10.1145/3240508.3240606

39. *Xiao T., Li S., Wang B., Lin L., Wang X.* End-to-end deep learning for person search // ArXiv, abs/1604.01850, 2016

40. *Layne R., Hospedales T.M., Gong S.* Investigating Open-World Person Re-identification Using a Drone // ECCV Workshops. 2014.
https://doi.org/10.1007/978-3-319-16199-0_16

41. *Fu D., Chen D., Bao J., Yang H., Yuan L., Zhang L., Li H., Chen D.* Unsupervised Pre-training for Person Re-identification // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 14745–14754.
https://doi.org/10.1109/CVPR46437.2021.01451

42. *Fabbri M., Brasó G., Maugeri G., Cetintas O., Gasparini R., Osep A., Calderara S., Leal-Taixe L., Cucchiara R.* MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. P. 10829–10839.
https://doi.org/10.1109/iccv48922.2021.01067

43. Makehuman community. Makehuman, 2020.
URL: http://www.makehumancommunity.org

44. Epic Games Incorporated. Unreal engine, 2020.
URL: https://www.unrealengine.com

45. *Barbosa I.B., Cristani M., Caputo B., Rognhaugen A., Theoharis T.* Looking beyond appearances: Synthetic training data for deep CNNs in re-identification // ArXiv, abs/1701.03153., 2018. https://doi.org/10.1016/j.cviu.2017.12.002

46. *Bak S., Carr P., Lalonde J.* Domain Adaptation through Synthesis for Unsupervised Person Re-identification // ECCV. ArXiv, abs/1804.10094, 2018.
https://doi.org/10.1007/978-3-030-01261-8_12

47. *Sun X., Zheng L.* Dissecting Person Re-Identification From the Viewpoint of Viewpoint // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 608–617. https://doi.org/10.1109/CVPR.2019.00070

48. *Wang Y., Liao S., Shao L.* Surpassing Real-World Source Training Data: Random 3D Characters for Generalizable Person Re-Identification // Proceedings of the 28th ACM International Conference on Multimedia. 2020. https://doi.org/10.1145/3394171.3413815

49. *Wang Y., Liang X., Liao S.* Cloning Outfits from Real-World Images to 3D Characters for Generalizable Person Re-Identification // ArXiv, abs/2204.02611. 2022. https://doi.org/10.48550/arXiv.2204.02611

50. Unity Technologies. 2020. Unity3D: Cross-platform game engine. URL: https://unity.com

51. *Zhong Z., Zheng L., Kang G., Li S., Yang Y.* Random Erasing Data Augmentation // AAAI. 2020. https://doi.org/10.1609/AAAI.V34I07.7000

52. *Ni X., Rahtu E.* FlipReID: Closing the Gap Between Training and Inference in Person Re-Identification // 2021 9th European Workshop on Visual Information Processing (EUVIP). 2021. P. 1–6. https://doi.org/10.1109/EUVIP50544.2021.9484010

53. *Li W., Xu F., Zhao J., Zheng R., Zou C., Wang M., Cheng Y.* HBReID: Harder Batch for Re-identification // ArXiv, abs/2112.04761, 2021. https://doi.org/10.48550/arXiv.2112.04761

54. *Huang Y., Zha Z., Fu X., Hong R., Li L.* Real-World Person Re-Identification via Degradation Invariance Learning // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 14072–14082. https://doi.org/10.1109/cvpr42600.2020.01409

55. *Jiang Y., Chen W., Sun X., Shi X., Wang F., Li H.* Exploring the Quality of GAN Generated Images for Person Re-Identification // Proceedings of the 29th ACM International Conference on Multimedia. 2021. https://doi.org/10.1145/3474085.3475547

56. *Wu C., Ge W., Wu A., Chang X.* Camera-Conditioned Stable Feature Generation for Isolated Camera Supervised Person Re-Identification // ArXiv, abs/2203.15210, 2022. https://doi.org/10.48550/arXiv.2203.15210

57. *Wang G., Lai J., Huang P., Xie X.* Spatial-Temporal Person Re-identification // ArXiv, abs/1812.03282. 2019. https://doi.org/10.1609/aaai.v33i01.33018933

58. *Yu Z., Jin Z., Wei L., Guo J., Huang J., Cai D., He X., Hua X.* Progressive Transfer Learning for Person Re-identification // IJCAI. 2019. https://doi.org/10.24963/ijcai.2019/586

59. *Sun Y., Zheng L., Yang Y., Tian Q., Wang S.* Beyond Part Models: Person Retrieval with Refined Part Pooling // ECCV. 2018. https://doi.org/10.1007/978-3-030-01225-0_30

60. *Bayoumi R.M., Hemayed E.E., Ragab M.E., Fayek M.B.* Person Re-Identification via Pyramid Multipart Features and Multi-Attention Framework // Big Data and Cognitive Computing. 2022. https://doi.org/10.3390/bdcc6010020

61. *Wang G., Yang S., Liu H., Wang Z., Yang Y., Wang S., Yu G., Zhou E., Sun J.* High-Order Information Matters: Learning Relation and Topology for Occluded

Person Re-Identification // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 6448–6457. https://doi.org/10.1109/CVPR42600.2020.00648

62. *Sun K., Xiao B., Liu D., Wang J.* Deep High-Resolution Representation Learning for Human Pose Estimation // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 5686–5696. https://doi.org/10.1109/CVPR.2019.00584
63. *Yang J., Zhang J., Yu F., Jiang X., Zhang M., Sun X., Chen Y., Zheng W.S.* Learning to Know Where to See: A Visibility-Aware Approach for Occluded Person Re-identification // Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021. P. 11885–11894.
64. *Fang H., Xie S., Tai Y., Lu C.* RMPE: Regional Multi-person Pose Estimation // IEEE International Conference on Computer Vision (ICCV). 2017. P. 2353–2362. https://doi.org/10.1109/ICCV.2017.256
65. *Chen X., Liu X., Liu W., Zhang X., Zhang Y., Mei T.* Explainable Person Re-Identification with Attribute-guided Metric Distillation // IEEE/CVF International Conference on Computer Vision (ICCV). 2022. P. 11793–11802. https://doi.org/10.1109/ICCV48922.2021.01160
66. *Dai Y., Sun Y., Liu J., Tong Z., Yang Y., Duan L.* Bridging the Source-to-target Gap for Cross-domain Person Re-Identification with Intermediate Domains // ArXiv, abs/2203.01682. 2022. https://doi.org/10.48550/arXiv.2203.01682
67. *Zhang H., Cisse M., Dauphin Y., Lopez-Paz D.* mixup: Beyond Empirical Risk Minimization // ArXiv, abs/1710.09412, 2018. https://doi.org/10.48550/arXiv.1710.09412
68. *Huang X., Belongie S.J.* Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization // 2017 IEEE International Conference on Computer Vision (ICCV). 2017. P. 1510–1519. https://doi.org/10.1109/ICCV.2017.167
69. *Avola D., Cascio M., Cinque L., Fagioli A., Petrioli C.* Person Re-Identification Through Wi-Fi Extracted Radio Biometric Signatures // IEEE Transactions on Information Forensics and Security. V. 17. 2022. P. 1145–1158. https://doi.org/10.1109/TIFS.2022.3158058
70. *Qi L., Shen J., Liu J., Shi Y., Geng X.* Label Distribution Learning for Generalizable Multi-source Person Re-identification // ArXiv, abs/2204.05903. 2022. https://doi.org/10.48550/arXiv.2204.05903
71. *Yang X., Zhou Z., Wang Q., Wang Z., Li X. Li H.* Cross-domain unsupervised pedestrian re-identification based on multi-view decomposition // Multimed Tools Appl. 2022. https://doi.org/10.1007/s11042-021-11797-w
72. *Elharrouss O., Almaadeed N., Al-Maadeed S.A., Bouridane A.* Gait recognition for person re-identification // J. Supercomput. 2021 No. 77. P. 3653–3672. https://doi.org/10.1007/s11227-020-03409-5
73. *Chao H., He Y., Zhang J., Feng J.* GaitSet: Regarding Gait as a Set for Cross-View Gait Recognition // ArXiv, abs/1811.06186, 2019. https://doi.org/10.1609/aaai.v33i01.33018126
74. *Jiang X., Qiao Y., Yan J., Li Q., Zheng W., Chen D.* SSN3D: Self-Separated Network to Align Parts for 3D Convolution in Video Person Re-Identification // Proceedings of the AAAI Conference on Artificial Intelligence. 2021. No. 35(2). P. 1691–1699. https://ojs.aaai.org/index.php/AAAI/article/view/16262

75. *Yang F., Wang X., Zhu X., Liang B., Li W.* Relation-based global-partial feature learning network for video-based person re-identification // Neurocomputing. 2022. V. 488. P. 424–435. https://doi.org/10.1016/j.neucom.2022.03.032.

76. *Lu Z., Zhang G., Huang G., Yu Z., Pun C., Ling K.* Video person re-identification using key frame screening with index and feature reorganization based on inter-frame relation // Int. J. Mach. Learn. Cyber. 2022. https://doi.org/10.1007/s13042-022-01560-4

77. *Yadav A., Vishwakarma D.K.* Person Re-Identification using Deep Learning Networks: A Systematic Review // ArXiv, abs/2012.13318. 2020. https://doi.org/10.48550/arXiv.2012.13318

78. *Zhang Z., Lan C., Zeng W., Jin X., Chen Z.* Relation-Aware Global Attention for Person Re-Identification // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020. P. 3183–3192. https://doi.org/10.1109/CVPR42600.2020.00325

79. *Pathak P., Eshratifar A.E., Gormish M.J.* Video Person Re-ID: Fantastic Techniques and Where to Find Them // AAAI. 2020. https://doi.org/10.1609/aaai.v34i10.7219

80. *Liu X., Zhang P., Yu C., Lu H., Yang X.* Watching You: Global-guided Reciprocal Learning for Video-based Person Re-identification // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 13329–13338. https://doi.org/10.1109/CVPR46437.2021.01313

81. *Gao S., Wang J., Lu H., Liu Z.* Pose-Guided Visible Part Matching for Occluded Person ReID // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 11741–11749. https://doi.org/10.1109/cvpr42600.2020.01176

82. *Zhang S., Yin Z., Wu X., Wang K., Zhou Q., Kang B.* FPB: Feature Pyramid Branch for Person Re-Identification // ArXiv, abs/2108.01901. 2021. https://doi.org/10.48550/arXiv.2108.01901

83. *Yang F., Li W., Liang. B., Han S., Zhu X.* Multi-stage attention network for video-based person re-identification // IET Comput. Vis. 2022. P. 1–11. https://doi.org/10.1049/cvi2.1210

84. *Wu G., Zhu X., Gong Sh.* Learning hybrid ranking representation for person re-identification // Pattern Recognition. V. 121. 2022. https://doi.org/10.1016/j.patcog.2021.108239

85. *Zhong Z., Zheng L., Cao D., Li S.* Re-ranking Person Re-identification with k-Reciprocal Encoding // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 3652–3661. https://doi.org/10.1109/CVPR.2017.389

86. *Bohush R.P., Ablameyko S.V. Adamovskiy E.R.* Image Similarity Estimation Based on Ratio and Distance Calculation between Features // Pattern Recognit. Image Anal. 2020. No. 30. P. 147–159. https://doi.org/10.1134/S1054661820020030

87. *He K., Zhang X., Ren S., Sun J.* Deep Residual Learning for Image Recognition // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. https://doi.org/10.1109/cvpr.2016.90

88. *Choi S., Kim T., Jeong M., Park H., Kim C.* Meta Batch-Instance Normalization for Generalizable Person Re-Identification // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 3424–3434. https://doi.org/10.1109/CVPR46437.2021.00343

89. *Huang G., Liu Z., Weinberger K.Q.* Densely Connected Convolutional Networks // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 2261–2269. https://doi.org/10.1109/CVPR.2017.243

90. *Chen P., Dai P., Liu J., Zheng F., Tian Q., Ji R.* Dual Distribution Alignment Network for Generalizable Person Re-Identification // AAAI. ArXiv, abs/2007.13249, 2021. https://doi.org/10.48550/arXiv.2007.13249

91. *Zhao C., Chen K., Wei Z., Chen Y., Miao D., Wang W.* Multilevel triplet deep learning model for person re-identification // Pattern Recognit. Lett. 2019. No. 117. P. 161–168. https://doi.org/10.1016/j.patrec.2018.04.029

92. *Yao Y., Jiang X., Fujita H., Fang Z.* A sparse graph wavelet convolution neural network for video-based person re-identification // Pattern Recognition. 2022. V. 129. https://doi.org/10.1016/j.patcog.2022.108708

93. *Lu P., Lu K., Wang W., Zhang J., Chen P., Wang B.* Real-Time Pedestrian Detection in Monitoring Scene Based on Head Model // Intelligent Computing Theories and Application. ICIC 2019. Lecture Notes in Computer Science. V. 11644. P. 558–568, Springer, Cham. https://doi.org/10.1007/978-3-030-26969-2_53

94. *Lee S., Kang Q., Madireddy S., Balaprakash P., Agrawal A., Choudhary A.N., Archibald R., Liao W.* Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time // 2019 IEEE International Conference on Big Data (Big Data). 2019. P. 830–839. https://doi.org/10.1109/BigData47090.2019.9006550

95. *Lewkowycz A.* How to decay your learning rate // ArXiv, abs/2103.12682, 2021. https://doi.org/10.48550/arXiv.2103.12682

96. *Lewkowycz A., Bahri Y., Dyer E., Sohl-Dickstein J., Gur-Ari G.* The large learning rate phase of deep learning: the catapult mechanism // ArXiv, abs/2003.02218, 2020. https://doi.org/10.48550/arXiv.2003.02218

97. *Ulyanov D., Vedaldi A., Lempitsky V.S.* Instance Normalization: The Missing Ingredient for Fast Stylization // ArXiv, abs/1607.08022, 2016. https://doi.org/10.48550/arXiv.1607.08022

98. *Chen H., Ihnatsyeva S., Bohush R., Ablameyko S.* Choice of activation function in convolution neural network in video surveillance systems // Programming and computer software. 2022. No. 5. P. 312–321. https://doi.org/10.1134/S0361768822050036

99. *Nair, Vinod, Geoffrey E. Hinton.* Rectified linear units improve restricted Boltzmann machines // ICML / 2010. P. 807–814.

100. *Maas Andrew L.* Rectifier non linearities improve neural network acoustic models // ICML. 2013. V. 30.

101. *Xu B., Wang N., Chen T., Li M.* Empirical Evaluation of Rectified Activations in Convolutional Network // ArXiv, abs/1505.00853, 2015. https://doi.org/10.48550/arXiv.1505.00853

102. *Clevert D., Unterthiner T., Hochreiter S.* Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs) // arXiv: abs/1511.07289v5, 2016. https://doi.org/10.48550/arXiv.1511.07289

103. *Klambauer G., Unterthiner T., Mayr A., Hochreiter S.* Self-Normalizing Neural Networks // ArXiv, abs/1706.02515, 2017. https://doi.org/10.48550/arXiv.1706.02515

104. *Hendrycks D., Gimpel K.* Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units. // ArXiv, abs/1606.08415, 2016. https://doi.org/10.48550/arXiv.1606.08415

105. *Ramachandran P., Zoph B., Le Q.V.* Swish: a Self-Gated Activation Function // arXiv: abs/1710.05941v2, 2017. https://doi.org/10.48550/arXiv.1710.05941

106. *Misra D.* Mish: A Self Regularized Non-Monotonic Neural Activation Function // ArXiv, abs/1908.08681, 2019. https://doi.org/10.48550/arXiv.1908.08681

107. *Lavi B., Ullah I., Fatan M., Rocha A.* Survey on Reliable Deep Learning-Based Person Re-Identification Models: Are We There Yet? // ArXiv, abs/2005.00355, 2020. https://doi.org/10.48550/arXiv.2005.00355

108. *Rao H., Miao C.* SimMC: Simple Masked Contrastive Learning of Skeleton Representations for Unsupervised Person Re-Identification // ArXiv, abs/ 2204.09826v1, 2022. https://doi.org/10.48550/arXiv.2204.09826

109. *Zheng Y., Zhou Y., Zhao J., Jian M., Yao R., Liu B., Chen Y.* A siamese pedestrian alignment network for person re-identification // Multim. Tools Appl. 2021. No. 80. P. 33951–33970. https://doi.org/10.1007/s11042-021-11302-3

110. *Zheng M., Karanam S., Wu Z., Radke R.J.* Re-Identification With Consistent Attentive Siamese Networks // IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 5728–5737. https://doi.org/10.1109/CVPR.2019.00588

111. *Hermans A., Beyer L., Leibe B.* In Defense of the Triplet Loss for Person Re-Identification // ArXiv, abs/1703.07737, 2017. https://doi.org/10.48550/arXiv.1703.07737

112. *Organisciak D., Riachy C., Aslam N., Shum H.* Triplet Loss with Channel Attention for Person Re-identification // J.WSCG. 2019. No. 27. https://doi.org/10.24132/JWSCG.2019.27.2.9

113. *Zhai Y., Guo X., Lu Y., Li H.* In Defense of the Classification Loss for Person Re-Identification // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2019. P. 1526–1535. https://doi.org/10.1109/CVPRW.2019.00194

114. *Alex D., Sami Z., Banerjee S., Panda S.* Cluster Loss for Person Re-Identification // Proceedings of the 11th Indian Conference on Computer Vision, Graphics and Image Processing. 2018. https://doi.org/10.1145/3293353.3293396

115. *Bai Z., Wang Z., Wang J., Hu D., Ding E.* Unsupervised Multi-Source Domain Adaptation for Person Re-Identification // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 12909–12918. https://doi.org/10.1109/CVPR46437.2021.01272

116. *Chen H., Lagadec B., Bremond F.* Unsupervised Lifelong Person Re-identification via Contrastive Rehearsal // ArXiv, abs/2203.06468, 2022. https://doi.org/10.48550/arXiv.2203.06468

117. *Zhang X., Li D., Wang Z., Wang J., Ding E., Shi J., Zhang Z., Wang J.* Implicit Sample Extension for Unsupervised Person Re-Identification // ArXiv, abs/2204.06892, 2022. https://doi.org/10.48550/arXiv.2204.06892

118. *Zhu K., Guo H., Yan T., Zhu Y., Wang J., Tang M.* Part-Aware Self-Supervised Pre-Training for Person Re-Identification // ArXiv, abs/2203.03931, 2022. https://doi.org/10.48550/arXiv.2203.03931

119. *Fu D., Chen D., Yang H., Bao J., Yuan L., Zhang L., Li H., Wen F., Chen D.* Large-Scale Pre-training for Person Re-identification with Noisy Labels // ArXiv, abs/2203.16533, 2022. https://doi.org/10.48550/arXiv.2203.16533

120. *Cho Y.H., Kim W.J., Hong S., Yoon S.* Part-based Pseudo Label Refinement for Unsupervised Person Re-identification // ArXiv, abs/2203.14675, 2022. https://doi.org/10.48550/arXiv.2203.14675

121. *Chen M., Wang Z., Zheng F.* Benchmarks for Corruption Invariant Person Re-identification // ArXiv, abs/2111.00880. 2021. https://doi.org/10.48550/arXiv.2111.00880

122. Dataset and Code. URL: https://www.pkuvmc.com/dataset.html

Статья представлена к публикации членом редколлегии О.П. Кузнецовым.

Поступила в редакцию 11.05.2022

После доработки 10.12.2022

Принята к публикации 26.01.2023