

УДК 004.932

DOI 10.52928/2070-1624-2022-38-4-13-25

ПРИНЦИПЫ ОРГАНИЗАЦИИ И АНАЛИЗ ПОДХОДОВ К ПОВЫШЕНИЮ ТОЧНОСТИ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛЮДЕЙ В РАСПРЕДЕЛЕННЫХ СИСТЕМАХ ВИДЕОНАБЛЮДЕНИЯ

С. А. ИГНАТЬЕВА

(Полоцкий государственный университет)

ORCID: <https://orcid.org/0000-0002-9780-5731>

Приведена классификация существующих систем повторной идентификации по таким критериям, как тип системы, количество и вид запросов, время работы. Рассмотрена общая схема, отражающая основной принцип работы систем повторной идентификации, а также основные подходы и методы для решения этой задачи с использованием сверточных нейронных сетей. Выполнено исследование существующих способов повышения точности работы алгоритмов и систем повторной идентификации. Проведен анализ влияния выбора гиперпараметров при обучении сверточных нейронных сетей на эффективность и динамику обучения алгоритма повторной идентификации.

Ключевые слова: повторная идентификация человека, сверточные нейронные сети, скорость обучения, размер пакета.

Введение. Повторная идентификация человека является актуальной задачей компьютерного зрения и представляет собой процесс идентификации человека в другом месте или в другое время на кадрах, полученных с нескольких камер видеонаблюдения. В общем случае в системе повторной идентификации имеется один или несколько запросов. Для каждого запроса выделяются признаки, которые затем сравниваются с признаками из имеющейся выборки изображений, называемой галереей, и принимается решение относительно того, является ли изображение из галереи искомым человеком. При этом приходится сталкиваться с рядом проблем, среди которых можно выделить изменчивость внешнего вида человека в зависимости от угла обзора, степени освещенности, различного разрешения камер; разнообразие положений, занимаемых человеком в пространстве; окклюзии; сложный фон; смещение доменов и ненадежность генерации ограничительных прямоугольников, содержащих изображения человека. Все это приводит к появлению большого числа методов и подходов, направленных на повышение устойчивости системы к различным негативным влияниям.

Широкая область применения систем повторной идентификации человека обуславливает существование большого количества алгоритмов и подходов для решения этой задачи, и, соответственно, различные способы классификации таких систем. Так, по типу используемых наборов данных можно выделить закрытые (Close-world) системы повторной идентификации, использующие готовые наборы данных для обучения и тестирования, и открытые системы (Open-world), в которых галерея изображений постоянно пополняется новыми кадрами [1]. 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, который изменяется с течением времени по мере поступления новых записей с камер наблюдения, ограничительные рамки необходимо генерировать в режиме реального времени, данные для обучения – аннотировать. Такие системы наиболее приближены к реальным условиям.

В зависимости от количества запросов [2] системы повторной идентификации можно разделить на повторную идентификацию одного человека и множества людей. В первом случае в галерее набора данных требуется найти человека по запросу, и задача повторной идентификации сводится к задаче поиска, или проверке, присутствует ли искомый человек в галерее. Во втором – для каждого человека устанавливается уникальный идентификатор, затем определяется, на каких кадрах каждый из этих людей встречается снова, и задача повторной идентификации сводится к задаче классификации [3].

По типу входных данных [1] системы повторной идентификации можно разделить на однородные и неоднородные. 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.

В зависимости от времени работы системы идентификацию разделяют на краткосрочную повторную и долгосрочную. Так, если каждый человек в наборе данных не меняет внешний вид с течением времени, т.е. все время 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 промежутка 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 [4].

Разнообразие областей применения и большое число проблем, с которыми приходится сталкиваться при решении задачи повторной идентификации, определяет существование большого числа алгоритмических решений, направленных на повышение скорости работы и точности повторной идентификации. Целью данного исследования является сравнительный анализ современных подходов в этой области.

1. Общая схема системы повторной идентификации человека на кадрах с нескольких камер видеонаблюдения

В общем виде любая система повторной идентификации подразумевает наличие нескольких камер видеонаблюдения K (рисунок 1). Все видеопоследовательности, полученные с камер, подаются на детектор, с помощью которого из отдельных кадров извлекаются ограничительные рамки, содержащие изображения людей, которые размещаются в галерее. Исследовательские наборы данных часто содержат уже извлеченные ограничительные рамки, и в таком случае можно считать, что процесс детектирования людей и формирование галереи выполнен заранее. Для каждого обнаруженного человека формируется дескриптор и помещается в таблицу признаков. Извлечение признаков в случае закрытой исследовательской системы выполняется предварительно, в случае открытой системы – в режиме реального времени. Большинство современных систем повторной идентификации для извлечения отличительных признаков человека используют сверточные нейронные сети (СНС). Это позволяет быстро и эффективно извлекать признаки на изображениях. Для каждого поступающего запроса также формируется дескриптор и выполняется поиск в таблице признаков путем ранжирования таблицы по установленным критериям подобия. Наилучшие совпадения выдаются как результат повторной идентификации. В зависимости от выбранного типа системы результатом повторной идентификации может быть ранжированный список наиболее совпадающих с запросом изображений или видеопоследовательность, на которой размечены идентифицированные лица.

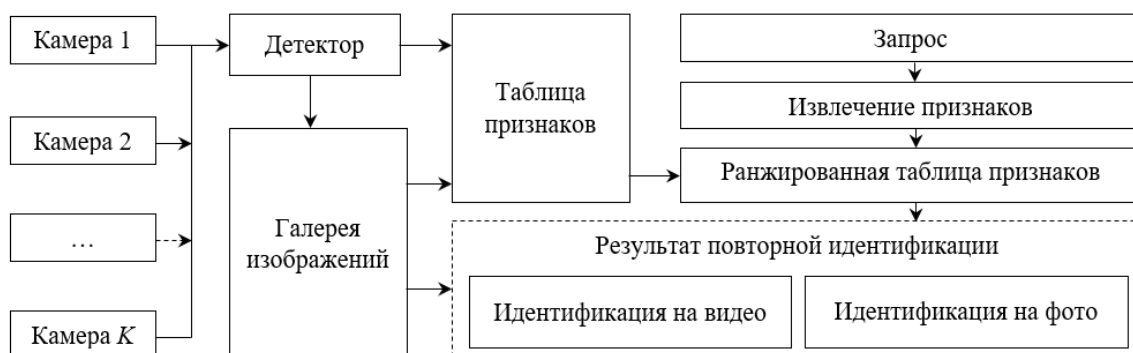


Рисунок 1. – Общая схема системы повторной идентификации человека

Для аннотированного набора данных $S = \{(I_1, y_1), (I_2, y_2), \dots, (I_n, y_n)\}$, где I_i – изображение, y_i – метка идентификатора, n – количество идентификаторов ($1 \leq i \leq n$), необходимо извлечь эффективный вектор признаков $f(\Theta; I_i)$, чтобы во всем пространстве признаков $\chi = \{x_i | x_i = f(\Theta; I_i), 1 \leq i \leq n\}$ расстояние между признаками для одинаковых идентификаторов было меньше, чем для людей с разными идентификаторами. Для этого необходимо стремиться к уменьшению ошибки E предсказания идентичности в S :

$$\min E(I_i, y_i) \in [y_i - g(f(I_i))],$$

где g – классификатор.

Качество извлеченных признаков зависит от распределения и разнообразия данных в S [5].

Использование СНС для извлечения признаков приводит к необходимости обучения модели, а для улучшения надежности извлекаемых признаков во время обучения и увеличения точности работы системы повторной идентификации уже обученной модели используются различные подходы.

При обучении для улучшения точности повторной идентификации можно подобрать оптимальные гиперпараметры, использовать увеличение обучающей выборки, аугментацию данных, подобрать наиболее эффективную функцию потерь, архитектуру СНС или рассматривать изображение не целиком, а разделяя его на фрагменты.

Для уже обученной модели улучшить работу алгоритма можно подбором наиболее эффективного способа ранжирования таблицы признаков, повторным ранжированием, учетом дополнительной информации о времени и месте съемки, атрибутах. Под атрибутами понимают семантическую информацию о человеке, имеющую значение для его идентификации: цвет и вид одежды, длину волос, сумки, рюкзаки, очки и другие значимые детали внешности.

2. Подходы для повышения точности систем повторной идентификации человека

2.1 Изменение гиперпараметров. На скорость и качество обучения СНС для извлечения признаков при решении любых задач компьютерного зрения, в том числе и задачи повторной идентификации, оказывают влияние выбранные перед началом обучения гиперпараметры, к которым можно отнести скорость обучения, ее изменение в процессе обучения, количество эпох, размер пакета, технические характеристики оборудования, на котором выполняют обучение.

Скорость обучения показывает, насколько изменяются весовые коэффициенты при каждом обновлении весов при обучении СНС. Существуют различные подходы, которые используют планировщики скорости обучения, позволяющие изменять скорость в процессе обучения по истечении определенного промежутка времени или по определенным критериям.

В [6] рассматривается механизм снижения скорости ADEL, предлагается отслеживать степень изменения весов и снижать скорость обучения каждый раз, когда веса перестают изменяться скачкообразно, т.е. когда изменения весов колеблются около локального минимума и снижение скорости обучения в этот момент позволяет обеспечить более быструю сходимость модели.

В [7] предлагается динамический механизм, состоящий из трех режимов изменения скорости обучения η , которые зависят от кривизны λ_0 поверхности функции потерь. Первая фаза изменения скорости – ленивая фаза, когда скорость обучения имеет относительно небольшое значение $\eta > \frac{2}{\lambda_0}$, из-за чего с шагом t изменение скорости остается практически постоянным в начале обучения. Вторая фаза – фаза катапульты, где скорость обучения принимает значения $\frac{2}{\lambda_0} < \eta < \eta_{\max}$. На этом этапе наблюдается экспоненциальный рост потерь и при этом быстрое уменьшение кривизны до тех пор, пока она не стабилизируется на значении $\lambda_{\text{final}} < \frac{\eta}{2}$. Как только выполняется это условие, достигается плоский минимум. И последняя фаза – фаза расхождения: когда скорость обучения превышает значение η_{\max} модель перестает обучаться. Кроме того, в [7] выдвигается предположение, которое затем подтверждается исследованиями, что использование больших скоростей обучения позволяет находить плоские минимумы, которые обобщаются лучше, чем резкие минимумы. К этому же, по мнению авторов, приводит и использование небольших пакетов для обучения.

Влияние размера пакета изображений за один проход по СНС исследуется в работе [8], где отмечается, что стремление к распараллеливанию вычислений обусловлено ускорением времени обучения модели и для наиболее эффективного ускорения необходимо увеличивать размер пакета. Однако это приводит к снижению сходимости модели, и в связи с этим размер пакета обычно выбирается в диапазоне от 16 до 64. В [8] предлагается стратегия, согласно которой в процессе обучения постепенно увеличивается размер минипакета и скорость обучения. Такой подход позволяет получить точность, близкую к точности с фиксированным небольшим размером пакета, при этом улучшая масштабируемость и снижая время обучения.

2.2 Формирование и увеличение обучающей выборки. На точность алгоритма повторной идентификации при обучении оказывают влияние размер, разнообразие и качество обучающей выборки. Формирование набора данных для обучения и тестирования – трудоемкий и дорогостоящий, с точки зрения оплаты труда, процесс. При этом следует помнить о существовании такой проблемы, как сдвиг домена [9; 10], когда наблюдается значительное снижение точности повторной идентификации при использовании системы в условиях, стилистически отличающихся от обучающей выборки. Частичным решением данной проблемы может быть объединение разных наборов данных, что рассматривается в [11; 12], в том числе и из необходимого домена [11; 13].

При использовании существующих наборов данных для обучения СНС, кроме проблемы сдвига домена, приходится сталкиваться с проблемой защиты персональных данных. Кроме того, некоторые наборы данных являются закрытыми: авторы предоставляют для исследований только извлеченные из изображений признаки [14], часть из которых можно использовать с ограничениями [15–17], т.е. при публикациях исследований авторы просят соблюдать конфиденциальность студентов, изображения которых использовались. Распространение подобных наборов данных возможно только при согласовании с авторами. Некоторые наборы данных (например, набор данных MTMC17 [18]) в настоящее время не доступен для исследований, DukeMTMC-ReID [19] был отозван из публичного доступа из-за нарушений гражданских прав,

прав человека и частной жизни студентов университета Duke, изображения которых использовались при формировании набора данных¹.

В связи с тем, что при создании набора данных необходимо явное согласие всех участников, некоторые исследователи для формирования обучающей выборки используют синтетически сгенерированные изображения. В [20] предлагается к рассмотрению 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 разместили точки обзора камеры, спланировали маршруты и перемещения пешеходов, установили параметры, связанные с поведением людей, характерным для людных мест. Использовались 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 набора данных Market1501 [21] и на 2,5% в метрике mAP при обучении на объединенном наборе данных Market1501 и CUHK03 [16].

В [5] рассматривается алгоритм генерации синтетических изображений для решения задачи адаптации домена. Для создания трехмерных реалистичных изображений людей используется MakeHuman², а для моделирования видеонаблюдения – платформа Unreal Engine 4 (UE4)³ 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 сгенерированных людях используются реальные изображения одежды, что отличает данный подход от существующих методов. При генерации намеренно добавляются люди с похожей внешностью и небольшими отличительными особенностями. Приводятся исследования, которые показывают, что использование сгенерированного набора данных показывает большую точность Rank1 при междоменном тестировании на наборе данных MTMC17, чем при использовании других синтетических наборов данных, таких как SOMAset [22], SyRI [23], PersonX [24], RandPerson [25]. Результаты подтверждаются при тестировании на Market1501 и DukeMTMC-ReID.

Главным преимуществом использования синтетических наборов данных является то, что при их формировании используется автоматическая генерация аннотаций, тогда как при формировании наборов данных из реальных изображений этот процесс обычно осуществляется вручную, что трудоемко, времязатратно и дорого.

Эффективным решением для повышения точности повторной идентификации является аугментация данных. Под аугментацией понимают увеличение обучающей выборки на основе только уже имеющихся данных. Самыми простыми методами аугментации являются различные преобразования изображений, такие как поворот, отражение, изменение размера, контраста, яркости, вариации цветовой составляющей, различная степень размытия и другие аналогичные манипуляции. Для повышения устойчивости к окклюзиям применяется метод «случайного стирания» [26], при котором прямоугольный фрагмент изображения, размер и форма которого выбираются случайным образом, заполняется нулевыми или случайными значениями (рисунок 2). Для задачи повторной идентификации тестирование этого метода аугментации осуществлялось на наборах данных Market1501, 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 примерно на 7% в метрике mAP. При тестировании на Market1501 и DukeMTMC-ReID результаты были скромнее: для разных алгоритмов повторной идентификации с различными СНС для извлечения признаков точность повторной идентификации Rank1 и mAP удавалось повысить от 1 до 4%.



Рисунок 2. – Примеры применения метода аугментации данных «Случайное стирание»

¹ Duke MTMC [Electronic resource]. URL: https://exposing.ai/duke_mtmc.

² Makehuman community. Makehuman, 2020 [Electronic resource]. URL: <http://www.makehumancommunity.org>.

³ Epic Games Incorporated. Unreal engine, 2020 [Electronic resource]. URL: <https://www.unrealengine.com>.

Более сложным методом аугментации данных является использование генеративно-сопоставительных сетей (Generative Adversarial network – GAN), которые используются для генерации правдоподобных изображений на основе уже имеющихся данных. Генеративно-сопоставительная сеть представляет собой алгоритм машинного обучения, в основе которого лежит комбинация двух нейронных сетей, одна из которых генерирует правдоподобные изображения, а другая пытается определить, являются ли изображения на ее входе подлинными. Применительно к задаче повторной идентификации использование GAN может быть направлено на улучшение способности извлечения эффективных признаков [27] или на решение проблем со смещением доменов [28].

В [27] рассматривается проблема, характерная для систем повторной идентификации, применяемых в реальных условиях, когда возможно присутствие различных факторов, ухудшающих качество полученных с камер видеонаблюдения изображений, таких как невысокий уровень освещения, низкое разрешение, влияние погодных условий. Так, например, если в момент наблюдения идет дождь, то система, обученная на данных, полученных при других условиях, не сможет с высокой точностью интерпретировать извлеченные признаки, и существует вероятность, что большое число извлеченных признаков будет учитывать сходства не между разными людьми, а между одинаковыми ухудшающими качество изображения факторами. Для решения этой проблемы необходимо изучить признаки различных влияний, ухудшающих качество изображений, однако это является сложной и некорректной задачей, так как в реальных условиях не может быть никаких аннотаций для описания этих факторов, а в обучающей выборке может не быть эталонных примеров. Для извлечения устойчивых к ухудшающим факторам изображений признаков авторы используют GAN для синтеза изображений с заранее известной степенью деградации.

В [28] 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 уже сгенерированными ранее изображениями, т.к. они могут снижать качество обучения, т.е. увеличивать время и при этом приводить к разбалансировке при обобщении. Для решения этой проблемы используется метод Local Outlier Factor (LOF), который контролирует плотность сгенерированных изображений, и в случае высокой плотности схожих сгенерированных изображений часть из них случайным образом отбрасывается. Такой подход позволяет не только повысить точность повторной идентификации, но и значительно повысить устойчивость системы к смещению домена. Так, в [28] приводятся результаты сравнения с другими алгоритмами, направленными на решение проблемы смещения домена. Предложенный подход позволяет получить точность повторной идентификации, соизмеримую с современными подходами.

2.3 Функции потерь. Процесс обучения нейронной сети для эффективного извлечения признаков заключается в корректировке весовых коэффициентов с целью уменьшения значения функции потерь L . Функция потерь отражает разницу между полученным результатом и ожидаемым, т.е. величину ошибки. Для задачи повторной идентификации наиболее распространенными являются кросс-энтропийная функция потерь (Cross-entropy loss) [10; 13; 29; 20] и триплетные потери (Triplet loss) [30; 31; 4; 32]. Отличительной чертой триплетных потерь является рассмотрение двух пар изображений: положительной ($y_a = y_p$), когда изображения принадлежат одному и тому же человеку, и отрицательной ($y_a \neq y_n$), когда два изображения принадлежат разным людям. Т.е. учитывается расстояние $d_{a,p}$ между признаками для положительной пары и расстояние $d_{a,n}$ между признаками разных людей. Чтобы сеть не только увеличивала расстояние между признаками разных людей, но и уменьшала расстояние для одинаковых, вводится коэффициент регуляризации m :

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

Для повышения точности повторной идентификации иногда используют несколько функций потерь. В [33] для определения наиболее эффективных признаков и наиболее значимых атрибутов предлагается две функции потерь: функция потерь метрического разделения L_d и функция потерь приоритетных атрибутов L_p :

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

где

$$L_d = \left| d_{i,j} - \sum_{k=1}^M d_{i,j}^k \right|$$

– функция потерь метрического разделения, основная задача которой состоит в том, чтобы разложить расстояние $d_{i,j}$, заданное целевой моделью, на вклады атрибутов в общий вектор признаков. $d_{i,j}^k$ – расстояние между

x_i и x_j для k -го из M атрибутов. Функция потерь приоритетных атрибутов L_p состоит из двух частей: L_{p1} – определяющей вклад общих атрибутов, и L_{p2} , определяющей вклад индивидуальных особенностей:

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

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

где $\hat{d}_{i,j} \approx \sum_{k=1}^M d_{i,j}^k$ – предсказанное значение расстояния между признаками;

M_E – количество уникальных атрибутов.

2.4 Подбор архитектуры СНС. В большинстве исследований, нацеленных на решение задачи повторной идентификации человека, при тестировании алгоритма используются различные архитектуры СНС, в числе которых встречается ResNet-50 в [11; 33; 34], DenseNet-121 [4; 29], MobileNetV2 [34]. Некоторые исследователи приводят результаты сравнения работы предлагаемых алгоритмов с использованием различных архитектур СНС для извлечения признаков. Так, в [4] выполняется сравнение ResNet-50 и DenseNet-121, где DenseNet-121 позволяет получить лучшие показатели точности в метриках Rank1 и mAP. В [33] оцениваются ResNet-34, ResNet-50 и ResNrt-101 и показывается, что увеличение глубины сети положительно сказывается на точности повторной идентификации. В [29] проводится сравнение точности работы для PCB [35], ResNet-50 и DenseNet-121. Анализ результатов этих экспериментов показывает, что наибольшей точностью в метриках Rank1 и mAP обладает PCB (Rank1 = 94,0, mAP = 82,8), затем идут DenseNet-121 (Rank1 = 90,8, mAP = 76,9) и ResNet-50 (Rank1 = 87,7, mAP = 72,2).

Кроме распространенных архитектур рассматриваются также их модификации: например, в [36] рассматривается влияние функции активации в СНС ResNet-50, DenseNet-121 на точность повторной идентификации. Функция активации влияет как на динамику обучения, так и на точность работы обученной модели. Результаты исследований подтверждают, что использованием вместо стандартной функции активации ReLU таких функций, как GeLU, Swish и Mish, можно повысить точность повторной идентификации. Дополнительные исследования показали, что применение этих функций увеличивает время обучения модели и при этом не позволяет получить достаточно стабильный результат. К числу предпочтительных функций активации для решения задачи повторной идентификации можно отнести GeLU и ReLU.

В [34] исследуется влияние способа нормализации данных на выходе сверточных слоев и предлагается MetaBIN (Meta Batch-Instance Normalization), которая использует комбинацию двух подходов: пакетную нормализацию и нормализацию экземпляров [37]. Пакетная нормализация позволяет получать информацию о различных стилях изображений в пакете, однако это может приводить к снижению точности повторной идентификации в невидимых доменах. Нормализация образцов отфильтровывает информацию о стиле, но вместе с тем может быть удалена и полезная информация. Для решения этих двух проблем вводится обучаемый параметр, который позволяет найти баланс между двумя подходами к нормализации и тем самым не только повысить точность повторной идентификации, но и сделать систему более устойчивой при работе в другом домене.

Для решения специфических задач, например, для неоднородных систем повторной идентификации, как в [38], где используются изображения с инфракрасной и камеры видимого диапазона, предлагается новая архитектура СНС MCLNet (Modality Confusion Learning Network). MCLNet основывается на частично-разделенной двухпоточной сети. Для повышения устойчивости системы к разнородным данным последовательно извлекаются признаки, характерные для каждой модальности по отдельности, а затем общие признаки. Так как видимые и инфракрасные образцы имеют разное распределение признаков и не могут быть согласованы для сравнения, сеть обучается игнорировать информацию о модальности и пытается извлекать общие представления для человека. Чтобы не упустить важные особенности разных людей, создается механизм запутывания обучения, в результате чего межмодальное несоответствие сводится к минимуму, а межмодальное сходство максимизируется.

В [4] также предлагается своя архитектура СНС RCSANet (Clothing Status Awareness Network), но для решения задачи долгосрочной повторной идентификации. Обычно методы, применяемые для решения подобной задачи, предполагают, что по прошествии определенного количества времени человек сменил одежду, прежде чем снова попасть в поле зрения камеры, однако такие подходы неэффективны, если в указанный промежуток времени человек не переоделся, и точность работы систем долгосрочной

повторной идентификации значительно снижается для таких людей. Для этого в [4] предлагается RCSANet, которая упорядочивает описание пешеходов и внедряет описание о состоянии одежды. RCSANet представляет собой двухпоточную систему, основанную на DenseNet-121, и включает ICE-поток (Inter-Class Enforcement), который позволяет максимизировать межклассовые различия для каждого человека, а также 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 позволил получить точность повторной идентификации в метрике Rank1 100% и 97,2% в случаях, когда смена одежды не осуществлялась, и 48% mAP, 50,2% Rank1, когда в тестовой выборке присутствовали люди, сменившие одежду.

2.5 Разделение изображения на фрагменты. Для повышения точности повторной идентификации некоторые исследователи предлагают рассматривать изображение человека не целиком, а по частям. Например, в [35] предлагается горизонтальное разделение изображения на 6 равных частей и изучение каждой части изображения по отдельности. Такой подход получил название Part-based Convolutional Baseline (PCB) и является надстройкой над СНС, деление выполняется не для самого изображения, а после первого сверточного слоя. Такой подход позволяет повысить точность повторной идентификации на 1–2% в метриках Rank1 и mAP. Недостатком является требование к расположению и содержанию каждой части, т.е. ЧЕЛОВЕК должен принимать строго вертикальное положение и части изображения должны располагаться в «правильных» местах. Ошибки генерации ограничительных рамок, когда часть человека оказывается обрезана, могут приводить к ошибкам идентификации.

В [39] рассматривается алгоритм повторной идентификации человека, основанный на рассмотрении человека по частям тела. Так, с помощью HR-Net [40] извлекаются ключевые точки человека, а затем исследуются признаки в окрестностях каждой ключевой точки. Данный подход направлен на уменьшение влияния окклюзий, и для решения этой проблемы при сопоставлении векторов признаков не учитываются признаки для тех ключевых точек, которые оказались скрыты.

В [32] рассматривается подход, схожий с двумя рассмотренными выше. Изображение разделяется на 6 горизонтальных частей, при этом для каждой из частей сеть пытается предсказать, есть ли на данном участке видимая часть человека. Если фрагмент изображения содержит видимые части человека, то с помощью оценщика поз AlphaPose [41] определяются ключевые точки человека и при предсказании, является ли обнаруженный человек искомым, признаки невидимых частей подавляются, что позволяет повысить точность повторной идентификации человека и увеличить устойчивость системы к окклюзиям.

2.6 Ранжирование признаков. Для поиска наилучших совпадений изображений галереи с запросом применяется ранжирование таблиц признаков, что представляет собой процесс пересортировки векторов признаков для всех изображений тестовой выборки таким образом, чтобы вверху таблицы находились векторы признаков, имеющие наибольшее сходство с дескриптором запроса. Для определения расстояния между вектором x_p изображения p запроса и вектором признаков x_{g_i} изображения g_i в галерее $G = \{g_i | i = 1, 2, \dots, N\}$ из N изображений могут использоваться подходы, представленные в таблице 1.

Таблица 1. – Методы определения расстояния между векторами признаков

$d(p, g_i) = \frac{x_p \cdot x_{g_i}}{\ x_p\ \ x_{g_i}\ }$	Косинусное расстояние (Cosine distance) [29; 13]
$d(p, g_i) = \ x_p - x_{g_i}\ _2^2$	Расстояние Евклида (Euclidean distance) [21; 4; 9; 12]
$d(p, g_i) = (x_p - x_{g_i})^T M^{-1} (x_p - x_{g_i}),$ где M – матрица ковариаций	Расстояние Махаланобиса (Mahalanobis distance) [42]
$d(p, g_i) = 1 - \frac{ R^*(p, k) \cap R^*(g_i, k) }{ R^*(p, k) \cup R^*(g_i, k) },$ где $R^*(p, k)$ и $R^*(g_i, k)$ – множества ближайших соседей	Расстояние Джакара для k -ближайших соседей (Jaccard distance) [42]

Для повышения точности повторной идентификации иногда используют повторное ранжирование, т.е. после первой сортировки по определенному алгоритму выполняется повторное ранжирование. Так, в [42] для первоначального ранжирования используется расстояние Махаланобиса. Из ранжированной таблицы выбираются первые k изображений из списка и включаются в множество кандидатов k -обратных ближайших соседей $R(p, k)$, после чего выполняется повторное ранжирование с использованием расстояния Джакара.

В [21] также применяется алгоритм повторного ранжирования. Сначала на основании расстояния Евклида выполняется первичное ранжирование векторов признаков, а затем из ранжированной таблицы $S(p, g)$ выбираются k первых результатов и для каждого из них выполняется поиск в галерее, в результате чего формируются новые ранжированные таблицы $S(r_i, g)$. Каждой новой таблице присваивается весовой коэффициент $\frac{1}{i+1}$, где $i=1, \dots, k$, и итоговая повторно ранжированная таблица признаков вычисляется как взвешенная сумма

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

2.7 Использование дополнительной информации. Еще одной возможностью увеличения точности повторной идентификации является использование дополнительной информации, которая предоставляется с набором данных в виде аннотаций или может быть извлечена автоматически при формировании ограничительных рамок: например, такая информация, как камера, с которой получено изображение (дает понимание о месте съемки) и номер кадра (позволяет получить сведения о времени съемки). Использование подобной информации предлагается в алгоритме повторной идентификации в [29]. Предложенный подход является двухпоточной системой, которая учитывает как визуальную составляющую, так и пространственно-временную. С помощью нейронной сети (DenseNet-121, ResNet-50 или PCB) извлекаются визуальные признаки объектов, пространственно-временная информация (т.е. номер камеры и номер кадра) содержится в названиях самих файлов. После ранжирования таблицы визуальных признаков из нее удаляются векторы признаков изображений, которые нерелевантны по пространственно-временным характеристикам, т.е. для тех людей, которые не могли в это время быть в этом месте. Для этого применяются гистограммы Парзена и логистическое сглаживание.

Алгоритмы повторной идентификации в большинстве случаев непрозрачны и приводят лишь к расстоянию между признаками людей. В [33] проводится исследование и предлагается подход, позволяющий определить и визуализировать признаки, которые система рассматривала при принятии решения о сходстве или различии людей. Данный подход дает ответы на вопросы о том, какие именно признаки для двух лиц были значимыми и какой вклад в этой разнице вносит каждый атрибут. Это достигается применением метода AMD (Attribute-guided 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, а также вводится функция потерь, которая позволяет сосредоточиться на характерных отличительных признаках. Эксперименты показывают, что такой подход позволяет не только визуализировать значимые признаки, но и способствует дополнительному улучшению точности повторной идентификации в целевых моделях. Приводятся также исследования, показывающие повышение точности повторной идентификации при тестировании алгоритма на междоменных данных.

3. Исследование влияния выбора гиперпараметров на точность реидентификации

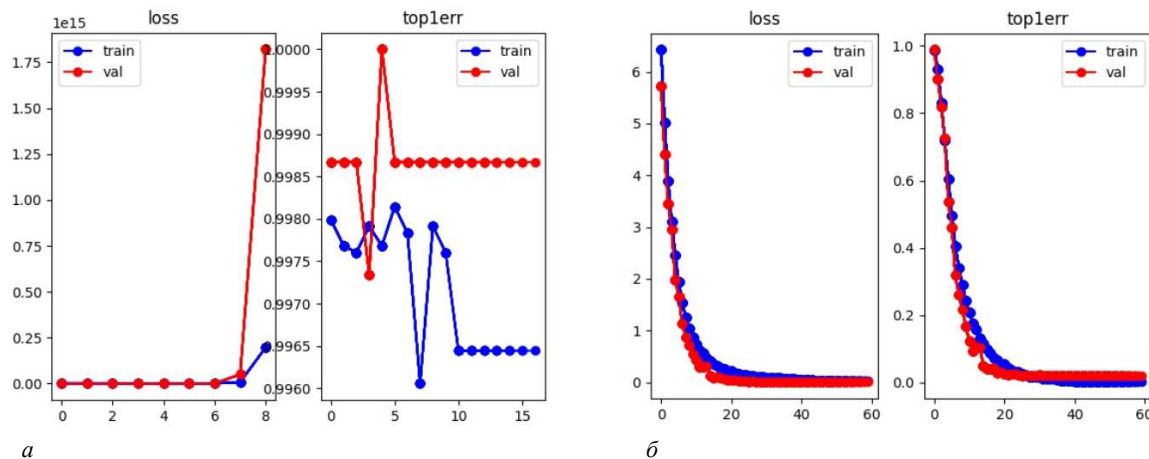
Гиперпараметры, выбранные перед началом обучения определяют динамику и качество обучения СНС. В таблице 2 представлены результаты обучения СНС ResNet-50 на наборе данных Market1501 с размером пакета 16 и различной скоростью обучения. Для эксперимента начальная скорость устанавливалась в интервале от 0,01 до 0,2 и с помощью планировщика скорости уменьшалась после 40-й эпохи в 0,01 раза. В качестве алгоритма повторной идентификации использовалась модель⁴, реализованная на фреймворке ruTorch, где в качестве функции потерь применялись кросс-энтропийные потери, а для ранжирования признаков использовалось косинусное расстояние. Обучение осуществлялось на персональном компьютере с основными характеристиками: Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb.

Таблица 2. – Влияние скорости обучения на точность повторной идентификации

Скорость обучения	Rank1	Rank5	Rank10	mAP	Время обучения
0,01	72,5059	88,6283	92,2830	46,8359	87 м 31 с
0,03	81,3242	92,4287	95,0416	57,6058	88 м 02 с
0,05	81,7696	92,3990	95,3682	59,5859	87 м 56 с
0,07	82,8979	93,2007	95,3385	59,7440	87 м 20 с
0,09	82,5713	92,7850	95,6354	60,1919	88 м 12 с
0,1	82,6990	93,0226	95,3682	61,7585	88 м 09 с

⁴ Person reID baseline PyTorch [Electronic resource]. URL: https://github.com/layumi/Person_reID_baseline_pytorch.

Результаты экспериментов показали, что увеличение скорости обучения позволяет повысить точность повторной идентификации и практически не влияет на время обучения модели. Однако высокая скорость обучения может приводить к взрывным градиентам, явлению, когда веса сети обновляются слишком быстро и значение функции потерь стремительно увеличивается, что и произошло в наших экспериментах при начальной скорости 0,2 (рисунок 3, а). Потери принимают значения порядка 10^{15} уже после 7-й эпохи обучения и далее продолжают увеличиваться, что приводит к невозможности дальнейшего обучения модели. Для сравнения приведены графики обучения при скорости 0,1 (рисунок 3, б), где потери и ошибка top-1 уменьшаются в процессе обучения.



а – график обучения со скоростью 0,2; б – график обучения со скоростью 0,1

Рисунок 3. – Графики обучения модели повторной идентификации с разной скоростью

Для исследования влияния размера пакета при обучении начальная скорость обучения равна 0,09. Несмотря на то, что при скорости 0,1 точность модели повторной идентификации выше, риск взрывных градиентов остается. Результаты экспериментов с разным размером пакета приведены в таблице 3.

Таблица 3. – Влияние размера пакета изображений при обучении на точность повторной идентификации

Размер пакета	Rank1	Rank5	Rank10	mAP	Время обучения
8	0,02969	0,08314	0,18409	0,02020	104 м 45 с
16	82,5713	92,7850	95,6354	60,1919	88 м 12 с
32	80,0178	91,7755	94,8634	56,9912	80 м 08 с
64	75,7423	90,4691	94,0024	51,0595	77 м 10 с

Из таблицы 3 видно, что увеличение размера пакета позволяет уменьшить время обучения, но снижает точность обученной модели. Выбор небольшого пакета для обучения не позволяет получить сети достаточную обобщающую способность, и приводит к тому, что точность повторной идентификации меньше 1%.

Таким образом, было установлено, что оптимальными параметрами для обучения на наборе данных Market1501 являются размер пакета 16 и скорость обучения 0,09.

Заключение. В результате анализа методов и подходов для повторной идентификации человека рассмотрены различные алгоритмические решения, направленные на повышение точности повторной идентификации и уменьшение влияния существующих проблем.

Для повышения устойчивости системы повторной идентификации к углам обзора камер, степени освещенности, схожести различных людей, вариативности внешнего вида одного и того же человека с разных ракурсов используются методы увеличения обучающей выборки и аугментация данных. Для решения проблемы смещения доменов применяются генеративно-состязательные сети (GAN), способы нормализации, синтетические наборы данных. Устойчивость к окклюзиям обеспечивается разделением изображений на фрагменты и применением методов аугментации. Комбинации различных подходов позволяют повышать точность повторной идентификации в целом.

При выборе гиперпараметров, функции потерь, функции активации и метода ранжирования признаков для определения наиболее эффективного подхода требуются дополнительные эмпирические исследования, т.к. при тестировании на разных наборах данных и СНС показатели точности могут отличаться.

Проведенные исследования влияния скорости обучения и размера пакета на точность повторной идентификации показали, что уменьшение размера пакета и увеличение скорости может повысить точность повторной идентификации, но при этом большая скорость обучения делает модель неустойчивой к взрывным градиентам, а небольшой размер пакета увеличивает время обучения и снижает обобщающую способность нейронной сети.

ЛИТЕРАТУРА

1. Deep Learning for Person Re-identification: A Survey and Outlook / M. Ye [et al.] // IEEE transactions on pattern analysis and machine intelligence. – 2021. – DOI: [10.1109/TPAMI.2021.3054775](https://doi.org/10.1109/TPAMI.2021.3054775).
2. Person Re-Identification across Data Distributions Based on General Purpose DNN Object Detector / R. Mihaescu [et al.] // Algorithms. – 2020. – № 13. – DOI: [10.3390/a13120343](https://doi.org/10.3390/a13120343).
3. Set-based classification for person re-identification utilizing mutual-information / H. Liu [et al.] // 2013 IEEE International Conference on Image Processing. – 2013. – P. 3078–3082. – DOI: [10.1109/ICIP.2013.6738634](https://doi.org/10.1109/ICIP.2013.6738634).
4. Clothing Status Awareness for Long-Term Person Re-Identification / Y. Huang [et al.] // 2021 IEEE/CVF International Conference on Computer Vision. – 2021. – P. 11895–11904. – DOI: [10.1109/ICCV48922.2021.01168](https://doi.org/10.1109/ICCV48922.2021.01168).
5. UnrealPerson: An Adaptive Pipeline towards Costless Person Re-identification / T. Zhang [et al.] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – P. 11501–11510. – DOI: [10.1109/CVPR46437.2021.01134](https://doi.org/10.1109/CVPR46437.2021.01134).
6. Lewkowycz, A. How to decay your learning rate / A. Lewkowycz // ArXiv, abs/2103.12682. – 2021. – DOI: [10.48550/arXiv.2103.12682](https://doi.org/10.48550/arXiv.2103.12682).
7. The large learning rate phase of deep learning: the catapult mechanism / A. Lewkowycz [et al.] // ArXiv, abs/2003.02218. – 2020. – DOI: [10.48550/arXiv.2003.02218](https://doi.org/10.48550/arXiv.2003.02218).
8. Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time / S. Lee [et al.] // 2019 IEEE International Conference on Big Data (Big Data). – 2019. – P. 830–839. – DOI: [10.1109/BigData47090.2019.9006550](https://doi.org/10.1109/BigData47090.2019.9006550).
9. Unsupervised Domain Adaptation with Noise Resistible Mutual-Training for Person Re-identification / F. Zhao [et al.] // ECCV. – 2020. – DOI: [10.1007/978-3-030-58621-8_31](https://doi.org/10.1007/978-3-030-58621-8_31).
10. Luo, C. Generalizing Person Re-Identification by Camera-Aware Invariance Learning and Cross-Domain Mixup / C. Luo, C. Song, Z. Zhang. // ECCV. – 2020. – DOI: [10.1007/978-3-030-58555-6_14](https://doi.org/10.1007/978-3-030-58555-6_14).
11. Style Normalization and Restitution for Generalizable Person Re-Identification / X. Jin [et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – P. 3140–3149. – DOI: [10.1109/cvpr42600.2020.00321](https://doi.org/10.1109/cvpr42600.2020.00321).
12. Generalizable Person Re-Identification by Domain-Invariant Mapping Network / J. Song [et al.] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 719–728. – DOI: [10.1109/CVPR.2019.00081](https://doi.org/10.1109/CVPR.2019.00081).
13. Ihnatsyeva, S. Joint Dataset for CNN-based Person Re-identification / S. Ihnatsyeva, R. Bohush, S. Ablameyko // 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.
14. Open-set Person Re-identification / S. Liao [et al.] // ArXiv, abs/1408.0872. – 2014. – DOI: [10.48550/arXiv.1408.0872](https://doi.org/10.48550/arXiv.1408.0872).
15. Li, W. Human Reidentification with Transferred Metric Learning. / W. Li, R. Zhao, X. Wang, // ACCV. – 2012. – DOI: [10.1007/978-3-642-37331-2_3](https://doi.org/10.1007/978-3-642-37331-2_3).
16. Li, W. Locally Aligned Feature Transforms across Views / W. Li, X. Wang // 2013 IEEE Conference on Computer Vision and Pattern Recognition. – 2013. – P. 3594–3601. – DOI: [10.1109/CVPR.2013.461](https://doi.org/10.1109/CVPR.2013.461).
17. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification / W. Li [et al.] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – P. 152–159. – DOI: [10.1109/CVPR.2014.27](https://doi.org/10.1109/CVPR.2014.27).
18. QPerson Transfer GAN to Bridge Domain Gap for Person Re-identification / L. Wei [et al.] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – P. 79–88. – 2018. – DOI: [10.1109/CVPR.2018.00016](https://doi.org/10.1109/CVPR.2018.00016).
19. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking / E. Ristani [et al.] // ArXiv, abs/1609.01775. – 2016. – DOI: [10.1007/978-3-319-48881-3_2](https://doi.org/10.1007/978-3-319-48881-3_2).
20. MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking? / M. Fabbri [et al.] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 10829–10839. – DOI: [10.1109/iccv48922.2021.01067](https://doi.org/10.1109/iccv48922.2021.01067).
21. Scalable Person Re-identification: A Benchmark. / L. Zheng [et al.] // 2015 IEEE International Conference on Computer Vision (ICCV). – 2015. – P. 1116–1124. – DOI: [10.1109/ICCV.2015.133](https://doi.org/10.1109/ICCV.2015.133).
22. Looking beyond appearances: Synthetic training data for deep CNNs in re-identification / I. B. Barbosa [et al.] // ArXiv, abs/1701.03153. – 2018. – DOI: [10.1016/i.cviu.2017.12.002](https://doi.org/10.1016/i.cviu.2017.12.002).
23. Bağ, S. Domain Adaptation through Synthesis for Unsupervised Person Re-identification / S. Bağ, P. Carr, J. Lalonde // ECCV. – 2018. – DOI: [10.1007/978-3-030-01261-8_12](https://doi.org/10.1007/978-3-030-01261-8_12).
24. Sun, X. Dissecting Person Re-Identification From the Viewpoint of Viewpoint / X. Sun, L. Zheng // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – P. 608–617. – 2019. – DOI: [10.1109/CVPR.2019.00070](https://doi.org/10.1109/CVPR.2019.00070).
25. Wang, Y. Surpassing Real-World Source Training Data: Random 3D Characters for Generalizable Person Re-Identification / Y. Wang, S. Liao, L. Shao // Proceedings of the 28th ACM International Conference on Multimedia. – 2020. – DOI: [10.1145/3394171.3413815](https://doi.org/10.1145/3394171.3413815).
26. Random Erasing Data Augmentation. / Z. Zhong [et al.] // AAAI. – 2020. – DOI: [10.1609/AAAI.V34I07.7000](https://doi.org/10.1609/AAAI.V34I07.7000).
27. Real-World Person Re-Identification via Degradation Invariance Learning. / Y. Huang [et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – P. 14072–14082. – 2020. – DOI: [10.1109/cvpr42600.2020.01409](https://doi.org/10.1109/cvpr42600.2020.01409).

28. Exploring the Quality of GAN Generated Images for Person Re-Identification / Y. Jiang [et al.] // Proceedings of the 29th ACM International Conference on Multimedia. – 2021. – DOI: [10.1145/3474085.3475547](https://doi.org/10.1145/3474085.3475547).
29. Spatial-Temporal Person Re-identification. / G. Wang [et al.] // ArXiv, abs/1812.03282. – 2019. – DOI: [10.1609/aaai.v33i01.33018933](https://doi.org/10.1609/aaai.v33i01.33018933).
30. Hermans, A. In Defense of the Triplet Loss for Person Re-Identification / A. Hermans, L. Beyrer, B. Leibe // ArXiv, abs/1703.07737. – 2017. – DOI: [10.48550/arXiv.1703.07737](https://doi.org/10.48550/arXiv.1703.07737).
31. Triplet Loss with Channel Attention for Person Re-identification / D. Organisciak [et al.] // J. WSCG. – No. 27. – 2019. – DOI: [10.24132/JWSCG.2019.27.2.9](https://doi.org/10.24132/JWSCG.2019.27.2.9).
32. Learning to Know Where to See: A Visibility-Aware Approach for Occluded Person Re-identification. / J. Yang [et al.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. – 2021. – P. 11885–11894. – DOI: [10.1109/ICCV48922.2021.01167](https://doi.org/10.1109/ICCV48922.2021.01167).
33. Explainable Person Re-Identification with Attribute-guided Metric Distillation / X. Chen [et al.] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 11793–11802. – DOI: [10.1109/ICCV48922.2021.01160](https://doi.org/10.1109/ICCV48922.2021.01160).
34. Meta Batch-Instance Normalization for Generalizable Person Re-Identification / S. Choi [et al.] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2021. – P. 3424–3434. – DOI: [10.1109/CVPR46437.2021.00343](https://doi.org/10.1109/CVPR46437.2021.00343).
35. Beyond Part Models: Person Retrieval with Refined Part Pooling / Y. Sun [et al.] // ECCV. – 2018. – DOI: [10.1007/978-3-030-01225-0_30](https://doi.org/10.1007/978-3-030-01225-0_30).
36. Игнатъева, С. А. Сравнительный анализ функций активации и их влияние на точность ре-идентификации людей с использованием сверточных нейронных сетей / С. А. Игнатъева // Современные проблемы математики и вычислительной техники : сб. материалов XII Респ. науч. конф. молодых ученых и студентов, Брест, 18–19 нояб. 2021 г. / Брест. гос. техн. ун-т. – Брест, 2021. – С. 44–48.
37. Ulyanov, D. Instance Normalization: The Missing Ingredient for Fast Stylization / D. Ulyanov, A. Vedaldi, V.S. Lempitsky // ArXiv, abs/1607.08022. – 2016. – DOI: [10.48550/arXiv.1607.08022](https://doi.org/10.48550/arXiv.1607.08022).
38. Cross-Modality Person Re-Identification via Modality Confusion and Center Aggregation. / X. Hao [et al.] // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 16383–16392. – DOI: [10.1109/ICCV48922.2021.0160](https://doi.org/10.1109/ICCV48922.2021.0160).
39. High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification / G. Wang [et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – P. 6448–6457. – DOI: [10.1109/CVPR42600.2020.00648](https://doi.org/10.1109/CVPR42600.2020.00648).
40. Deep High-Resolution Representation Learning for Human Pose Estimation / K. Sun [et al.] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2019. – P. 5686–5696. – DOI: [10.1109/CVPR.2019.00584](https://doi.org/10.1109/CVPR.2019.00584).
41. RMPE: Regional Multi-person Pose Estimation / H. Fang [et al.] // 2017 IEEE International Conference on Computer Vision (ICCV). – 2017. – P. 2353–2362. – DOI: [10.1109/ICCV.2017.256](https://doi.org/10.1109/ICCV.2017.256).
42. Re-ranking Person Re-identification with k-Reciprocal Encoding / Z. Zhong [et al.] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 3652–3661. – DOI: [10.1109/CVPR.2017.389](https://doi.org/10.1109/CVPR.2017.389).

REFERENCES

1. Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., & Hoi, S.C. (2021). Deep Learning for Person Re-identification: A Survey and Outlook. *IEEE transactions on pattern analysis and machine intelligence*, PP. DOI: [10.1109/TPAMI.2021.3054775](https://doi.org/10.1109/TPAMI.2021.3054775).
2. Mihaescu, R., Chindea, M., Paleologu, C., Carata, S., & Ghenescu, M. (2020). Person Re-Identification across Data Distributions Based on General Purpose DNN Object Detector. *Algorithms*, 13(12), 343. DOI: [10.3390/a13120343](https://doi.org/10.3390/a13120343).
3. Liu, H., Qin, L., Cheng, Z., & Huang, Q. (2013). Set-based classification for person re-identification utilizing mutual-information. *2013 IEEE International Conference on Image Processing* (3078–3082). DOI: [10.1109/ICIP15918.2013](https://doi.org/10.1109/ICIP15918.2013).
4. Huang, Y., Wu, Q., Zhong, Y., & Zhang, Z. (2021). Clothing Status Awareness for Long-Term Person Re-Identification. *2021 IEEE/CVF International Conference on Computer Vision* (11895–11904). DOI: [10.1109/ICCV48922.2021.01168](https://doi.org/10.1109/ICCV48922.2021.01168).
5. Zhang, T., Xie, L., Wei, L., Zhuang, Z., Zhang, Y., Li, B., & Tian, Q. (2021). UnrealPerson: An Adaptive Pipeline towards Costless Person Re-identification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (11501–11510). DOI: [10.1109/CVPR46437.2021.01134](https://doi.org/10.1109/CVPR46437.2021.01134).
6. Lewkowycz, A. (2021). How to decay your learning rate. *ArXiv, abs/2103.12682*. DOI: [10.48550/arXiv.2103.12682](https://doi.org/10.48550/arXiv.2103.12682).
7. Lewkowycz, A., Bahri, Y., Dyer, E., Sohl-Dickstein, J., & Gur-Ari, G. (2020). The large learning rate phase of deep learning: the catapult mechanism. *ArXiv, abs/2003.02218*. DOI: [10.48550/arXiv.2003.02218](https://doi.org/10.48550/arXiv.2003.02218).
8. Lee, S., Kang, Q., Madireddy, S., Balaprakash, P., Agrawal, A., Choudhary, A.N., ... Liao, W. (2019). Improving Scalability of Parallel CNN Training by Adjusting Mini-Batch Size at Run-Time. *2019 IEEE International Conference on Big Data (Big Data)* (830–839). DOI: [10.1109/BigData47090.2019.9006550](https://doi.org/10.1109/BigData47090.2019.9006550).
9. Zhao, F., Liao, S., Xie, G., Zhao, J., Zhang, K., & Shao, L. (2020). Unsupervised Domain Adaptation with Noise Resistant Mutual-Training for Person Re-identification. *ECCV*. DOI: [10.1007/978-3-030-58621-8_31](https://doi.org/10.1007/978-3-030-58621-8_31).
10. Luo, C., Song, C., & Zhang, Z. (2020). Generalizing Person Re-Identification by Camera-Aware Invariance Learning and Cross-Domain Mixup. *ECCV*. DOI: [10.1007/978-3-030-58555-6_14](https://doi.org/10.1007/978-3-030-58555-6_14).
11. Jin, X., Lan, C., Zeng, W., Chen, Z., & Zhang, L. (2020). Style Normalization and Restitution for Generalizable Person Re-Identification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (3140–3149). DOI: [10.1109/cvpr42600.2020.00321](https://doi.org/10.1109/cvpr42600.2020.00321).
12. Song, J., Yang, Y., Song, Y., Xiang, T., & Hospedales, T.M. (2019). Generalizable Person Re-Identification by Domain-Invariant Mapping Network. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (719–728). DOI: [10.1109/CVPR.2019.00081](https://doi.org/10.1109/CVPR.2019.00081).

13. Ihnatsyeva, S., Bohush, R., & Ablameyko, S. (2021). Joint Dataset for CNN-based Person Re-identification. *Pattern Recognition and Information Processing (PRIP'2021) Proceedings of the 15th International Conference* (33–37). Minsk: United Institute of Informatics Problems of the National Academy of Sciences of Belarus.
14. Liao, S., Mo, Z., Hu, Y., & Li, S. (2014). Open-set Person Re-identification. *ArXiv, abs/1408.0872*. DOI: [10.48550/arXiv.1408.0872](https://doi.org/10.48550/arXiv.1408.0872).
15. Li, W., Zhao, R., & Wang, X. (2012). Human Reidentification with Transferred Metric Learning. *ACCV*. DOI: [10.1007/978-3-642-37331-2_3](https://doi.org/10.1007/978-3-642-37331-2_3).
16. Li, W., & Wang, X. (2013). Locally Aligned Feature Transforms across Views. *2013 IEEE Conference on Computer Vision and Pattern Recognition* (3594–3601). DOI: [10.1109/CVPR.2013.461](https://doi.org/10.1109/CVPR.2013.461).
17. Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). DeepReID: Deep Filter Pairing Neural Network for Person Re-identification. *2014 IEEE Conference on Computer Vision and Pattern Recognition* (152–159). DOI: [10.1109/CVPR.2014.27](https://doi.org/10.1109/CVPR.2014.27).
18. Wei, L., Zhang, S., Gao, W., & Tian, Q. (2018). Person Transfer GAN to Bridge Domain Gap for Person Re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (79–88). DOI: [10.1109/CVPR.2018.00016](https://doi.org/10.1109/CVPR.2018.00016).
19. Ristani, E., Solera, F., Zou, R.S., Cucchiara, R., & Tomasi, C. (2016). Performance Measures and a Data Set for Multi-target, Multi-camera Tracking. *ArXiv, abs/1609.01775*. DOI: [10.1007/978-3-319-48881-3_2](https://doi.org/10.1007/978-3-319-48881-3_2).
20. Fabbri, M., Brasó, G., Maugeri, G., Cetintas, O., Gasparini, R., Osep, A., ... Cucchiara, R. (2021). MOTSynth: How Can Synthetic Data Help Pedestrian Detection and Tracking? *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (10829–10839). DOI: [10.1109/iccv48922.2021.01067](https://doi.org/10.1109/iccv48922.2021.01067).
21. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., & Tian, Q. (2015). Scalable Person Re-identification: A Benchmark. *2015 IEEE International Conference on Computer Vision (ICCV)* (1116–1124). DOI: [10.1109/ICCV.2015.133](https://doi.org/10.1109/ICCV.2015.133).
22. Barbosa, I. B., Cristani, M., Caputo, B., Roghhaugen, A., & Theoharis, T. (2018). Looking beyond appearances: Synthetic training data for deep CNNs in re-identification. *ArXiv, abs/1701.03153*. DOI: [10.1016/j.cviu.2017.12.002](https://doi.org/10.1016/j.cviu.2017.12.002).
23. Bık, S., Carr, P., & Lalonde, J. (2018). Domain Adaptation through Synthesis for Unsupervised Person Re-identification. *ECCV*. DOI: [10.1007/978-3-030-01261-8_12](https://doi.org/10.1007/978-3-030-01261-8_12).
24. Sun, X., & Zheng, L. (2019). Dissecting Person Re-Identification From the Viewpoint of Viewpoint. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (608–617). DOI: [10.1109/CVPR.2019.00070](https://doi.org/10.1109/CVPR.2019.00070).
25. Wang, Y., Liao, S., & Shao, L. (2020). Surpassing Real-World Source Training Data: Random 3D Characters for Generalizable Person Re-Identification. *Proceedings of the 28th ACM International Conference on Multimedia*. DOI: [10.1145/3394171.3413815](https://doi.org/10.1145/3394171.3413815).
26. Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. (2020). Random Erasing Data Augmentation. *AAAI*. DOI: [10.1609/AAAI.V34I07.7000](https://doi.org/10.1609/AAAI.V34I07.7000).
27. Huang, Y., Zha, Z., Fu, X., Hong, R., & Li, L. (2020). Real-World Person Re-Identification via Degradation Invariance Learning. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (14072–14082). DOI: [10.1109/cvpr42600.2020.01409](https://doi.org/10.1109/cvpr42600.2020.01409).
28. Jiang, Y., Chen, W., Sun, X., Shi, X., Wang, F., & Li, H. (2021). Exploring the Quality of GAN Generated Images for Person Re-Identification. *Proceedings of the 29th ACM International Conference on Multimedia*. DOI: [10.1145/3474085.3475547](https://doi.org/10.1145/3474085.3475547).
29. Wang, G., Lai, J., Huang, P., & Xie, X. (2019). Spatial-Temporal Person Re-identification. *ArXiv, abs/1812.03282*. DOI: [10.1609/aaai.v33i01.33018933](https://doi.org/10.1609/aaai.v33i01.33018933).
30. Hermans, A., Beyer, L., & Leibe, B. (2017). In Defense of the Triplet Loss for Person Re-Identification. *ArXiv, abs/1703.07737*. DOI: [10.48550/arXiv.1703.07737](https://doi.org/10.48550/arXiv.1703.07737).
31. Organisciak, D., Riachy, C., Aslam, N., & Shum, H. (2019). Triplet Loss with Channel Attention for Person Re-identification. *J. WSCG*, 27. DOI: [10.24132/JWSCG.2019.27.2.9](https://doi.org/10.24132/JWSCG.2019.27.2.9).
32. Yang, J., Zhang, J., Yu, F., Jiang, X., Zhang, M., Sun, X., ... Zheng, W. S., (2021) 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* (11885–11894). DOI: [10.1109/ICCV48922.2021.01167](https://doi.org/10.1109/ICCV48922.2021.01167).
33. Chen, X., Liu, X., Liu, W., Zhang, X., Zhang, Y., & Mei, T. (2021). Explainable Person Re-Identification with Attribute-guided Metric Distillation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (11793–11802). DOI: [10.1109/ICCV48922.2021.01160](https://doi.org/10.1109/ICCV48922.2021.01160).
34. Choi, S., Kim, T., Jeong, M., Park, H., & Kim, C. (2021). Meta Batch-Instance Normalization for Generalizable Person Re-Identification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (3424–3434). DOI: [10.1109/CVPR46437.2021.00343](https://doi.org/10.1109/CVPR46437.2021.00343).
35. Sun, Y., Zheng, L., Yang, Y., Tian, Q., & Wang, S. (2018). Beyond Part Models: Person Retrieval with Refined Part Pooling. *ECCV*. DOI: [10.1007/978-3-030-01225-0_30](https://doi.org/10.1007/978-3-030-01225-0_30).
36. Ignat'eva, S. A. (2021). Sravnitel'nyj analiz funkcij aktivacii i ih vliyanie na tochnost' re-identifikacii lyudej s ispol'zovaniem svertochnyh nejronnyh setej [Comparative activation functions analysis and their impact on the person re-identification accuracy using convolutional neural networks], *Sovremennye problemy matematiki i vychislitel'noj tekhniki [Modern problems of mathematics and computer technology]* (44–48). Brest: BrSTU (In Russ.).
37. Ulyanov, D., Vedaldi, A., & Lempitsky, V. S. (2016). Instance Normalization: The Missing Ingredient for Fast Stylization. *ArXiv, abs/1607.08022*. DOI: [10.48550/arXiv.1607.08022](https://doi.org/10.48550/arXiv.1607.08022).
38. Hao, X., Zhao, S., Ye, M., & Shen, J. (2021). Cross-Modality Person Re-Identification via Modality Confusion and Center Aggregation. *2021 IEEE/CVF International Conference on Computer Vision (ICCV)* (16383–16392). DOI: [10.1109/ICCV48922.2021.01160](https://doi.org/10.1109/ICCV48922.2021.01160).
39. Wang, G., Yang, S., Liu, H., Wang, Z., Yang, Y., Wang, S., ... Sun, J. (2020). High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* ((6448–6457). DOI: [10.1109/CVPR42600.2020.00648](https://doi.org/10.1109/CVPR42600.2020.00648).

40. Sun, K., Xiao, B., Liu, D., & Wang, J. (2019). Deep High-Resolution Representation Learning for Human Pose Estimation. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (5686–5696). DOI: [10.1109/CVPR.2019.00584](https://doi.org/10.1109/CVPR.2019.00584).
41. Fang, H., Xie, S., Tai, Y., & Lu, C. (2017). RMPE: Regional Multi-person Pose Estimation. *2017 IEEE International Conference on Computer Vision (ICCV)* (2353–2362). DOI: [10.1109/ICCV.2017.256](https://doi.org/10.1109/ICCV.2017.256).
42. Zhong, Z., Zheng, L., Cao, D., & Li, S. (2017). Re-ranking Person Re-identification with k-Reciprocal Encoding. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (3652–3661). DOI: [10.1109/CVPR.2017.389](https://doi.org/10.1109/CVPR.2017.389).

Поступила 31.03.2022

ORGANIZATION PRINCIPLES AND APPROACHES ANALYSIS TO IMPROVING THE PERSON RE-IDENTIFICATION ACCURACY IN DISTRIBUTED VIDEO SURVEILLANCE SYSTEMS

S. IHNATSYEVA

The paper presents a classification of existing re-identification systems according to such criteria as system type, requests number and type, and operating time. The general scheme is discussed, which reflects the basic operation principle of re-identification systems, and the main approaches and methods for solving this problem using convolutional neural networks are considered. The study ways existing to improve re-identification algorithms and systems accuracy has been carried out. The influence analysis hyperparameters choice in convolutional neural networks training on the efficiency and dynamics re-identification algorithm training is carried out.

Keywords: person re-identification, convolutional neural networks, learning rate, batch size.