

УДК 004.932

DOI 10.52928/2070-1624-2025-45-2-9-17

**АЛГОРИТМИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ
ДЛЯ ПОВТОРНОЙ ИДЕНТИФИКАЦИИ ЛЮДЕЙ
НА ОСНОВЕ НЕЙРОСЕТЕВЫХ ПРИЗНАКОВ ИЗОБРАЖЕНИЙ ЛИЦА И ФИГУРЫ**

**канд. техн. наук С. А. ИГНАТЬЕВА, д-р техн. наук, доц. Р. П. БОГУШ, Н. А. ТОМАШЕВИЧ
(Полоцкий государственный университет имени Евфросинии Полоцкой)**

**проф. Х. ЧЕН
(университет Чжэцзян Шурен, Китай)**

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

R. P. Bohush ORCID: <https://orcid.org/0000-0002-6609-5810>

M. A. Tamashevich ORCID: <https://orcid.org/0009-0009-9214-5103>

H. Chen ORCID: <https://orcid.org/0000-0003-4229-4505>

Рассматриваются основные проблемы реализации открытых систем повторной идентификации человека на основе нейросетевых технологий. Исследуется влияние гиперпараметров при обучении сверточной нейронной сети на динамику обучения и точность алгоритма реидентификации человека. Экспериментальный подбор гиперпараметров нейронной сети для реидентификации человека состоит из двух этапов. На первом этапе необходимо определить наиболее эффективную скорость обучения и размер пакета изображений. На втором этапе подбирается количество эпох обучения с учетом изменения размера пакета и скорости в процессе обучения нейронной сети. Осуществлен ряд экспериментов с применением ResNet-50 и DenseNet-121 на базах данных PolReID1077, Market-1501, DukeMTMC-ReID и MSMT17. Определены такие гиперпараметры, как размер пакета, скорость обучения и количество эпох обучения нейронной сети. Результаты экспериментов подтвердили повышение точности реидентификации. Кроме того, время обучения нейронных сетей с применением предложенного метода корректировки гиперпараметров позволяет сократить время обучения по сравнению с использованием метода обучения на базовой модели. Представлен алгоритм для повторной идентификации людей, использующий один глобальный и два локальных нейросетевых дескриптора для описания черт человека на основе нейросетевых признаков его лица и фигуры. Изображение фигуры человека описывается вектором из 1536 элементов, полученным с применением DenseNet-121. Предложенный подход обеспечивает высокую точность повторной идентификации как по изображению лица, так и в более сложных условиях, когда лицо скрыто или видимы лишь фрагменты фигуры. Представлены результаты экспериментов. Описана программная реализация прототипа для повторной идентификации человека в открытом мире.

Ключевые слова: реидентификация, сверточные нейронные сети, гиперпараметры, скорость обучения, размер пакета.

Введение. Повторная идентификация представляет собой задачу компьютерного зрения, предполагающую определение присутствия человека, ранее идентифицированного системой видеонаблюдения в одном и том же месте в разное время или в разных местах видеонаблюдения по запросу. В качестве запроса обычно используется изображение человека. Для сопоставления запроса с имеющимися в базе изображениями людей (галерее) необходимо сформировать дескриптор, отражающий наиболее значимые отличительные черты. Эффективным подходом для формирования дескриптора изображений является применение сверточных нейронных сетей (СНС). Для исследовательских целей обычно рассматриваются закрытые (close-world) системы реидентификации. В таких системах количество изображений или видеопоследовательностей фиксировано и не изменяется с течением времени. Перед использованием набор данных состоит из ограниченного числа видеопоследовательностей или изображений, полученных с нескольких камер видеонаблюдения. Данные в таких наборах аннотируются и подготавливаются заранее, для каждого запроса в галерее присутствуют изображения соответствующего человека. Открытые системы (open-world) повторной идентификации не предполагают предварительной подготовки и аннотирования данных. Галерея изображений формируется по мере поступления видеопоследовательностей с камер наблюдения и детектирования на них людей. Такие системы приближены к реальным условиям, что сопряжено с рядом дополнительных затруднений. Так, в [1] рассматриваются современные подходы к реализации открытых систем повторной идентификации и отмечается, что подобные системы совмещают две задачи компьютерного зрения: обнаружение и повторную идентификацию. Это означает необходимость совместной реализации данных задач. В [2] основной проблемой открытых систем повторной идентификации называется сходство посторонних людей с объектом запроса. Для решения этой проблемы применяется состязательное обучение, при котором для каждого запроса генерируются синтетические изображения, схожие с целевым объектом. В [3; 4] поднимается проблема изменчивости внешности людей с течением времени. В качестве решения рассматривается анализ действий людей (например, походки [5]) или трехмерное моделирование фигуры человека [4]. В [3] авторы рассматривают вариации освещенности как одну из ключевых проблем повторной идентификации и предлагают

алгоритм, позволяющий восстановить уровень освещенности изображений, полученных при различных условиях видеонаблюдения, до уровня тестовых данных.

В данной работе поднимается сразу несколько проблем: перекрытие лиц другими людьми и фоновыми объектами; схожесть внешнего вида разных людей, а также неизбежные изменения внешности отдельного человека с течением времени; необходимость формирования набора данных для обучения сверточной нейронной сети (СНС) в условиях, аналогичных условиям развертывания системы повторной идентификации [6]. В работе описаны способы решения указанных проблем и представлена программная реализация прототипа системы повторной идентификации человека в открытом мире.

1. Анализ существующих проблем

Широко распространенной проблемой нейросетевых моделей для реидентификации людей является проблема смещения домена (domain shift), которая возникает в ситуации, когда обучение и тестирование осуществляются на данных из разных наборов [7; 8]. Под доменом понимают комплект изображений, которые были получены при одинаковых условиях в одной мультикамерной системе видеонаблюдения. Каждое изображение в наборе данных зависит от совокупности факторов, включающих разрешение камер, фон, условия освещения и даже внешний вид людей, т. е. статистически европейцы будут иметь вид, отличный от азиатов, летняя одежда будет отличаться от зимней и т. д. Система, обученная на наборе данных, полученном с внутренних камер видеонаблюдения, может иметь низкую точность на тестовой выборке, содержащей изображения людей, снятые наружными камерами видеонаблюдения. Алгоритмы, направленные на решение такого рода проблем, относят к междоменной повторной идентификации (Cross-domain ReID). Другой особенностью, в значительной мере определяющей точность реидентификации человека, является цвет одежды. Однако цветопередача различных видеокамер и освещенность сцены видеонаблюдения могут значительно исказить цветовые характеристики одного и того же объекта. Помимо разного уровня освещенности сцены видеонаблюдения, при повторной идентификации возникает проблема выделения общих отличительных признаков для человека, изображения которого получены с разных ракурсов. Схожей проблемой является вариативность поз, которые может принимать человек. При этом, кроме сложности выделения общих признаков, в ситуациях, когда человек наклоняется, стоит или сидит, возникает проблема искажения восприятия телосложения при нормализации изображений до единого размера $H \times W$, где H – высота изображения; W – ширина.

Значимой проблемой является разность характеристик камер видеонаблюдения и степень удаленности объекта от них, так как при нормализации изображений людей до единого размера они будут иметь разные качественные характеристики: детализацию, резкость, размытие. Применение при видеонаблюдении камер с различными характеристиками оказывает влияние как на разрешение полученных изображений, так и на цветопередачу. Извлечение отличительных признаков человека может быть затруднено фоновыми помехами. На одном изображении могут присутствовать два разных человека, и возникает необходимость определить, чей идентификатор следует присвоить изображению. Цвет одежды, волос или атрибутов может сливаться с фоном, что приводит к утрате или неверной интерпретации важных характерных особенностей. По многим причинам разные люди могут иметь схожий внешний вид (например, одинаковую униформу и прическу), но отличаться малозаметной деталью, например, обувью. При этом один и тот же человек может в течение короткого промежутка времени изменить наличие какого-либо атрибута, что приведет к изменению набора признаков. Это сопоставимо с проблемой разности условий видеонаблюдения, таких как ракурс съемки, удаленность от камеры, ее характеристики и т. д. Как и во многих задачах компьютерного зрения, при повторной идентификации возникает проблема окклюзий, когда между камерой видеонаблюдения и объектом интереса расположен посторонний объект: часть тела другого человека, предмет интерьера или элемент ландшафта. Перекрытие изображения человека другими объектами вносит в общий дескриптор дополнительные признаки, не относящиеся к отличительным особенностям интересующей личности, и одновременно затрудняет выделение значимых черт.

При рассмотрении открытых систем повторной идентификации существенной проблемой являются ошибки детектора, формирующего ограничительные прямоугольники с изображением обнаруженных людей. К таким ошибкам можно отнести примеры изображений (рисунок 1), в которых части тел оказываются обрезаны (рисунок 1, а–в) или когда в качестве изображения людей детектируются портреты (рисунок 1, г), статуи (рисунок 1, д), отдельные части тела человека (рисунок 1, е, ж).



Рисунок 1. – Примеры изображений людей, демонстрирующие ошибки детектора

Для всех нейросетевых моделей, в том числе решающих задачи компьютерного зрения, характерна проблема переобучения. Она возникает при недостаточном количестве или разнообразии обучающих данных и характеризуется их запоминанием. Вследствие этого модель теряет способность к обобщению на новых изображениях. Одним из способов решения проблемы переобучения является регуляризация. В машинном обучении под регуляризацией понимают добавление ограничений к архитектуре нейронной сети, функции потерь или обучающим наборам данных [9]. Примерами регуляризации являются: прореживание нейронных связей СНС [10]; использование разных функций активации [11]; L1- и L2-регуляризация¹; аугментация данных, при которой часть изображения может быть удалена или замещена другой информацией [9; 12; 13]. Аугментация применяется для увеличения разнообразия обучающей выборки на основе имеющихся данных [13] за счет преобразований изображений.

При разработке системы повторной идентификации сложно заранее учесть и исключить все потенциальные проблемы, поэтому применяемые методы, подходы и алгоритмы обычно направлены на решение одной или нескольких конкретных задач. На основе проведенного анализа сформулированы основные проблемы, а также предложены способы уменьшения их влияния на точность повторной идентификации (таблица 1).

Таблица 1. – Проблемы, оказывающие влияние на точность реидентификации и подходы, направленные на их решение

Проблема	Подход к решению проблемы
Смещение домена	– увеличение обучающей выборки за счет данных из различных доменов; – обучение СНС извлечению и исключению признаков, характерных для условий видеонаблюдения; – поиск способов пакетной нормализации для повышения точности при использовании данных, принадлежащих разным доменам; – перенос обучения (transfer learning); – расширение обучающей выборки путем применения GAN (Generative Adversarial Networks) для переноса стилистики изображений, характерной для разных доменов; – алгоритм формирования и обработки дескриптора
Уровень освещенности	– формирование разнообразной обучающей выборки; – аугментация данных; – генерация с помощью GAN синтетических изображений, моделирующих вариации освещения и погодных условий; – алгоритм формирования и обработки дескриптора
Характеристики камеры	– применение для обучения СНС изображений, полученных с большого числа камер видеонаблюдения с разнообразными характеристиками; – алгоритм формирования и обработки дескриптора
Вариабельность поз	– разнообразная обучающая выборка, включающая соответствующие примеры; – аугментация обучающих данных; – алгоритм формирования и обработки дескриптора
Сложный фон	– обучающая выборка; – аугментация данных; – алгоритм формирования и обработки дескриптора
Внешний вид людей	– обучающая выборка; – аугментация данных; – алгоритм формирования и обработки дескриптора
Окклюзии	– акцентирование внимания на локальных признаках; – аугментация данных; – обучающая выборка; – алгоритм формирования и обработки дескриптора
Переобучение СНС	– увеличение обучающей выборки; – регуляризация; – аугментация; – перенос обучения (использование предварительно обученных моделей); – регулирование гиперпараметров при обучении
Ошибки детектора	– выбор средства детектирования людей; – качество обучения детектора; – настройка параметров срабатывания детектора

Анализ таблицы 1 показывает, что подходы к повышению точности повторной идентификации людей можно разделить на две группы: применяемые на этапе обучения СНС и используемые на этапе постобработки извлеченных признаков. На этапе обучения точность повторной идентификации может быть повышена за счет выбора архитектуры СНС, использования наиболее эффективных значений гиперпараметров,

¹ L2 and L1 Regularization in Machine Learning // AnalyticsSteps [Электронный ресурс]. – URL: <https://www.analytics-steps.com/blogs/l2-and-l1-regularization-machine-learning> (дата обращения: 26.09.2025).

увеличения объема и разнообразия обучающих данных. На этапе постобработки извлеченных признаков точность повторной идентификации будет зависеть от алгоритма, определяющего тип признаков.

2. Повышение точности повторной идентификации человека на этапе обучения СНС

2.1. Увеличение обучающей выборки. Увеличение количества обучающих примеров для тренировки СНС позволяет уменьшить негативное влияние на точность повторной идентификации таких факторов, как смещение домена, уровень освещенности, различие характеристик камер видеонаблюдения, вариативность поз, внешность людей, сложный фон. Сбор и аннотирование собственного набора данных – долгий и трудоемкий процесс, поэтому для достижения удовлетворительной точности работы системы количество обучающих примеров, относящихся к домену разворачиваемой системы, можно увеличить за счет готовых наборов данных. Для обучения СНС в работе использовался набор данных PolReID1077², сформированный в условиях умеренного климата. Он включает изображения людей, полученные в разное время года и суток, при отличающихся погодных условиях, с внутренних и наружных камер видеонаблюдения, как при естественном, так и при искусственном освещении. Объем обучающей выборки увеличен за счет существующих наборов данных, таких как Market-1501 [14], DukeMTMC-ReID [15], MSMT17 [16], CUHK02 [17] и CUHK03 [18]. В общей сложности обучающий набор включает 109 772 изображения для 6381 человека.

2.2. Расширение тренировочного набора изображений и двухэтапное обучение СНС. Для увеличения разнообразия тренировочной выборки без увеличения количества изображений используется метод аугментации, включающий циклический сдвиг и исключение цветности, применяемые к отдельным изображениям, а замещение фрагмента уменьшенным изображением применяется к пакету, подаваемому на вход СНС [19]. При таком методе аугментации на одном изображении присутствуют два разных человека, каждый со своим идентификатором, что позволяет повысить устойчивость к переобучению, окклюзиям, изменениям освещенности и разрешения обрабатываемых изображений, а также к зависимости от местоположения признаков отдельных фрагментов фигуры человека. Такие изображения значительно отличаются от реальных, которые могут быть получены с камер видеонаблюдения. Поэтому не смотря на увеличение точности при таком методе аугментации значения функции потерь при обучении остаются на высоком уровне. Для снижения этого показателя применяется технология двухэтапного обучения, при которой на первом этапе СНС обучается на аугментированных изображениях, а затем выполняется точная настройка на исходных изображениях, что позволяет уменьшить значения функции потерь и повысить точность реидентификации.

2.3. Настройка гиперпараметров. В работе рассматривалось влияние на точность повторной идентификации таких гиперпараметров, как количество эпох обучения, размер пакета, скорость обучения. Количество эпох обучения определяет, сколько раз будут обновляться весовые коэффициенты нейронной сети. Скорость обучения характеризует величину изменения весовых коэффициентов на каждой итерации. Размер пакета отражает количество изображений из набора данных, которые одновременно подаются на вход сети для обработки. Настройка гиперпараметров осуществлялась в несколько этапов. Сначала определялись стартовые значения, при которых точность повторной идентификации максимальна. Исследования проводились на наборе данных Market-1501 для двух СНС: DenseNet-121 [20] и ResNet-50 [21]. Для пакетов размерами $b = 2, 4, 8, 16, 32$ и 64 изображения обучение осуществлялось со скоростью $\eta \in [0,001; 0,009]_{h=0,001}$; $\eta \in [0,01; 0,09]_{h=0,01}$; $\eta \in [0,1; 0,4]_{h=0,1}$ для выходного слоя. Скорость скрытых слоев устанавливалась в 10 раз меньше. После 40-й эпохи скорость для всех слоев уменьшалась в 10 раз. Точность обученной модели оценивалась каждые 5 эпох по трем метрикам: Rank1, mAP и mINP. Rank1 отражает количество запросов, для которых верное решение имеет наибольшее сходство с запросом. Метрика mAP представляет собой усредненное значение средних точностей для каждого запроса, а mINP позволяет оценить сложность поиска наиболее проблемного предсказания, показывая, насколько далеко в ранжированном списке предсказаний оказывается последнее изображение человека, соответствующее запросу. Лучшие результаты получены для СНС ResNet-50 при размере пакета $b = 8$ и скорости обучения $\eta = 0,03; 0,04; 0,05$, а также при $b = 16$ и $\eta = 0,07$.

Второй этап экспериментальных исследований направлен на определение варианта настройки гиперпараметров для оценки возможности дальнейшего повышения точности повторной идентификации. Он включает несколько серий экспериментов. Исследования проводились с использованием СНС ResNet-50 и DenseNet-121. Для обучения использовался набор данных Market-1501.

Для *первой серии* экспериментов размер пакета удваивался каждые 20 эпох начиная с $b = 8$. Скорость обучения изменялась, как и для базовой модели, каждые 40 эпох. Рассматривались значения $\eta_0 = 0,04$; $\eta_0 = 0,05$; $\eta_0 = 0,06$; $\eta_0 = 0,07$; $\eta_{i:40} = 0,1\eta_{i-40}$. Лучший результат для выбранных значений, использованных в каждом варианте эксперимента, получен при $\eta_0 = 0,06$ и равен по метрике Rank1 = 82,0071; mAP = 60,3425; mINP = 23,9121.

Во *второй серии* экспериментов количество изображений удваивалось каждые 30 эпох, начиная с $b = 16$, при тех же значениях скоростей обучения. Лучшие показатели получены при скорости обучения $\eta_0 = 0,05$ и составили: Rank1 = 80,2850; mAP = 56,8707; mINP = 20,0197.

² Dataset PolReID1077 [Электронный ресурс] // GitHub. – URL: <https://github.com/SvetlanaIgn/PolReID1077> (дата обращения: 26.09.2025).

В *третьей серии* экспериментальных исследований размер пакета первые десять эпох составлял 32 изображения, с 10 по 20 эпохи – 16, с 20 по 60 – 8 изображений, а затем – снова 16. Для экспериментов рассматривались значения скорости обучения $\eta_0 = 0,04$; $\eta_0 = 0,05$; $\eta_0 = 0,06$; $\eta_i = k \cdot \eta_{i-1}$, $k \in [0,9; 0,99]_{h=0,01}$. Лучшие показатели реидентификации получены при $\eta_0 = 0,05$ и $k = 0,99$. При этом Rank1 = 84,7684; mAP = 64,5680; mINP = 28,1097.

Случайным образом размер пакета изменялся в *четвертой серии* экспериментов, а скорость обучения корректировалась так же, как и на предыдущем этапе. Лучшие показатели точности получены при $k = 0,96$, однако значения по всем метрикам оценки точности меньше предыдущего этапа и Rank1 = 83,7686; mAP = 61,5846; mINP = 24,6478.

Пятая серия экспериментов аналогична второй и отличается только стартовым значением размера пакета $b = 8$. Лучшие показатели достигаются при скорости обучения $\eta_0 = 0,05$: Rank1 = 83,3135; mAP = 62,4157; mINP=25,3371.

Для *шестой серии* экспериментов размер пакета удваивался каждые 30 эпох, начиная с $b = 8$, а скорость обучения уменьшалась вдвое каждые 20 эпох для $\eta_0 = 0,04$; $\eta_0 = 0,05$; $\eta_0 = 0,06$. Точность реидентификации при $\eta_0 = 0,05$ в этой серии была максимальной: Rank1 = 84,3527; mAP = 62,6090; mINP = 25,5956.

В *седьмой серии* экспериментов размер пакета удваивался после 40-й и 60-й эпох, а скорость обучения уменьшалась каждые 10 эпох. В данной серии экспериментов наилучшая точность составила: по метрике Rank1 = 83,7589; по метрике mAP = 63,0670; по метрике mINP = 25,9548.

Наилучшие результаты с точки зрения точности повторной идентификации были получены для метода настройки гиперпараметров в восьмой экспериментальной серии. Размер пакета с 1-й по 60-ю эпоху составлял 8 изображений, а с 60-й до 90-й использовался пакет из 16 изображений. После эпохи 90 размер пакета составляет 32 изображения. Скорость обучения варьируется от $\eta_0 = 0,05$ до η_i согласно формуле

$$\eta_i = 0,98 \eta_{i-1},$$

где i – номер эпохи.

Точность реидентификации при таком способе корректировки гиперпараметров составила: по метрике Rank1 = 85,4810; по метрике mAP = 65,4131; по метрике mINP = 28,5910.

2.4. Алгоритм повторной идентификации с использованием составного дескриптора изображений фигуры человека и признаков лица. Для решения проблем перекрытия лиц, вариативности внешнего вида одного человека и одинаковой одежды у разных людей предлагается алгоритм повторной идентификации, включающий формирование дескриптора черт лица человека и составного вектора признаков, описывающих его фигуру. Для формирования дескриптора лица человека используется обученная сверточная нейронная сеть ResNet50@WebFace600K от Insightface³, формирующая 512 отличительных признаков.

На первом этапе на кадрах видеопоследовательности выполняется обнаружение людей и обрамление каждого из них ограничительным прямоугольником. Для каждого из обнаруженных людей выполняется детектирование ключевых точек фигуры человека. Любое изображение с менее чем тремя обнаруженными ключевыми точками считается невалидным и исключается из дальнейшей обработки. Для всех корректных изображений формируется составной нейросетевой дескриптор, описывающий всю фигуру человека, его верхнюю и нижнюю части. Верхняя часть фигуры будет считаться видимой, если хотя бы одна ключевая точка обнаружена на любой из следующих четырех частей тела: лице, плече, локте, запястье, бедре. Нижняя часть фигуры человека видна, если обнаружена хотя бы одна ключевая точка на двух частях тела, например, на бедрах, коленях или лодыжках. Если человек на изображении смотрит в камеру, для него формируется вектор признаков, характеризующий лицо.

Таким образом, алгоритм выполняет следующие шаги для каждого кадра видеопоследовательности: для каждого человека в кадре с помощью сверточной нейронной сети YOLOv7⁴ создаются ограничивающий прямоугольник и выделяются ключевые точки; определяются ограничивающие прямоугольники для верхней и нижней половины тела человека на основе обнаруженных ключевых точек; формирование признаков локализованных областей выполняется с использованием сверточной нейронной сети DenseNet-121; полученные признаки представляются в виде вектора F со следующими элементами: F_p (глобальный дескриптор), F_T (дескриптор верхней части), F_B (дескриптор нижней части); на основе анализа ключевых точек A_1 (нос), A_2 (левый глаз), A_3 (правый глаз), A_4 (левое ухо), A_5 (правое ухо) вычисляется приблизительная область лица, что приводит к снижению вычислительных затрат при последующей обработке; точная локализация ограничительного прямоугольника лица осуществляется на основе CHC buffalo_l; идентификатор группы изображений для одного человека назначается вектору

³ InsightFace.Ai [Электронный ресурс]. – URL: <https://www.insightface.ai> (дата обращения: 26.09.2025).

⁴ Official YOLOv7 [Электронный ресурс] // Github. – URL: <https://github.com/airhorns/yolov7-pose> (дата обращения: 26.09.2025).

признаков, при этом используется древовидная структура, чтобы избежать создания новых групп и присвоения идентификаторов изображениям с признаками, схожими с уже существующей группой; выполняется группировка людей на основе признаков лиц, если они были сформированы; возвращается уникальный идентификатор, основанный на признаках лица текущего человека; если признаки лица человека были извлечены и им была назначена группа, возвращается идентификатор (номер группы), к которому они были добавлены; присвоение различных идентификаторов также выполняется для схожих векторов признаков, если они получены для изображений людей, полученных с пространственно разнесенных камер с непересекающимися областями обзора или на одном кадре одновременно, т. к. один и тот же человек не может быть запечатлен на нескольких камерах одновременно в разных местах или несколько раз на изображении, полученном с одной камеры.

3. Экспериментальные исследования

При проведении экспериментов в качестве экстрактора признаков используется сеть DenseNet-121, обучение которой осуществляется на обучающих наборах PolReID1077 (сгенерирован в условиях видеонаблюдения, аналогичных условиям тестирования разработанной системы повторной идентификации), Market-1501, DukeMTMC-ReID, MSMT17 и Occluded-Duke [22], обучающая выборка которого аналогична DukeMTMC-ReID, а тестовая сформирована с искусственными перекрытиями изображений людей, добавленными в случайных местах. Обучение проводилось на персональном компьютере с процессором Intel Core i5 3,11 ГГц, 16 ГБ оперативной памяти и видеокартой Nvidia GeForce RTX-3060 6 ГБ. Функция потерь рассчитывается, а весовые коэффициенты корректируются, как описано в [23]. Приведены результаты точности обучения нейронных сетей с фиксированным размером пакета и скоростью обучения, изменяемой каждые 40 эпох, а также с предлагаемой корректировкой гиперпараметров (см. подраздел 2.3) в процессе обучения: на различных наборах данных для базового алгоритма повторной идентификации⁵; на расширенном обучающем наборе с использованием анализа локальных признаков, как описано в подразделе 2.4, двухэтапного обучения и аугментации данных. Определение точности сверточной нейронной сети предлагается выполнять с использованием метрик Rank1, mAP и mINP. Результаты экспериментов приведены в таблице 2. Представленные значения оценки реидентификации по различным метрикам показывают, что использование корректировки гиперпараметров в процессе обучения позволяет повысить точность реидентификации как относительно базовой модели для различных наборов данных, так и при использовании расширенного обучающего набора, двухэтапной технологии обучения на дополненных данных и алгоритма с составным дескриптором нейронной сети. Наибольший прирост метрики Rank1 относительно базового алгоритма составил 39,16 для набора данных Occluded-DukeMTMC. По метрике mAP максимальное улучшение точности также отмечено для набора данных Occluded-Duke и составило 38,48. Наибольший прирост mINP достигнут для наборов данных Market-1501 и PolReID1077 и составил 33,97 и 33,31 соответственно.

Таблица 2. – Точность реидентификации людей с различными параметрами обучения

Алгоритм и параметры обучения		Базовая модель						Расширенная обучающая выборка, технология двухэтапного обучения на аугментированных данных, анализ локальных и глобальных признаков					
		Стандартный планировщик скорости обучения			Предлагаемая корректировка гиперпараметров			Стандартный планировщик скорости обучения			Предлагаемая корректировка гиперпараметров		
Метрики		Rank1	mAP	mINP	Rank1	mAP	mINP	Rank1	mAP	mINP	Rank1	mAP	mINP
Набор данных	M*	84,02	64,29	28,15	85,48	65,43	29,75	94,12	85,73	61,41	94,89	86,47	62,12
	D**	72,80	53,16	15,88	75,00	55,14	17,27	87,79	76,68	38,70	87,93	77,12	38,82
	P***	90,33	68,39	23,43	92,35	70,60	26,04	97,19	89,32	56,53	97,29	90,14	56,74
	MS****	54,98	26,67	02,61	57,11	27,88	03,55	80,32	57,55	13,05	80,83	57,68	13,94
	OD*****	32,31	21,50	04,67	35,06	23,63	05,45	69,55	58,17	27,86	71,47	59,98	28,83

* – набор данных Market-1501.

** – набор данных DukeMCMC-ReID.

*** – набор данных PolReID1077.

**** – набор данных MSMT17.

***** – набор данных Occluded-Duke.

⁵ Pytorch ReID [Электронный ресурс] // GitHub. – URL: https://github.com/layumi/Person_reID_baseline_pytorch (дата обращения: 26.09.2025).

4. Программная реализация открытой системы повторной идентификации

На рисунке 2 показан главный экран реализованной системы повторной идентификации человека. В нижней части экрана отображаются все видеозаписи, доступные системе. Выбор видеозаписей для отображения в левой части экрана осуществляется нажатием левой кнопки мыши. Обнаруженные изображения людей сортируются по группам в соответствии с присвоенными им идентификаторами и отображаются в правой части экрана.

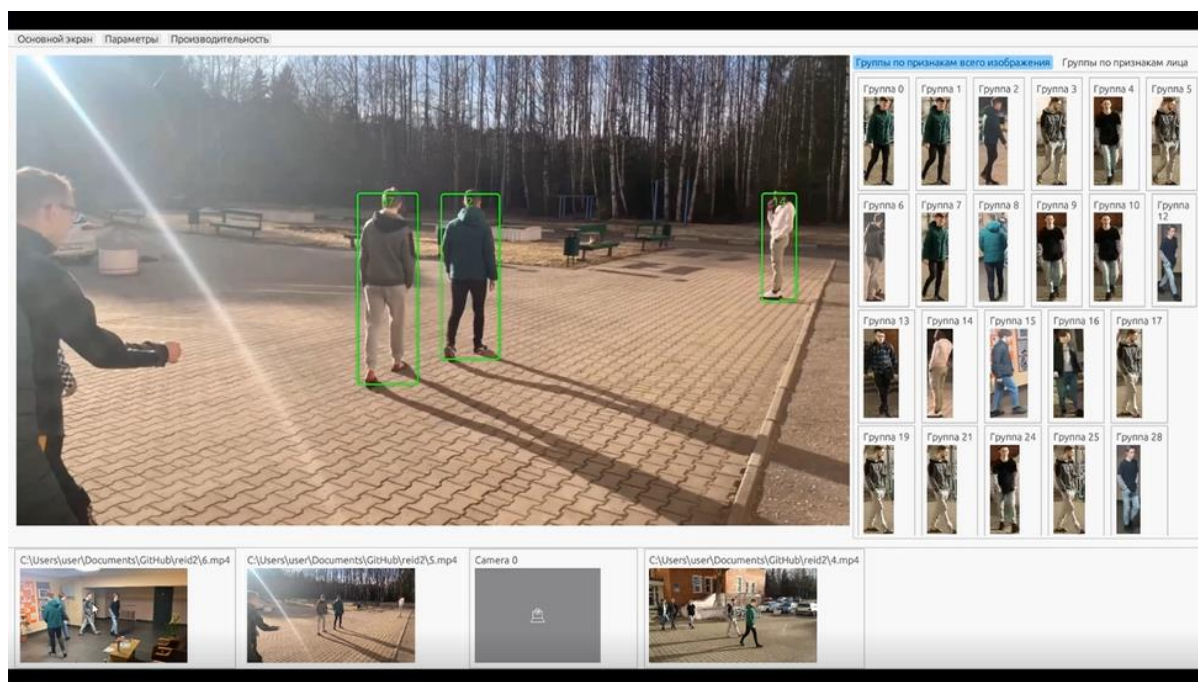


Рисунок 2. – Главный экран пользовательского интерфейса открытой системы повторной идентификации

Первоначальное распределение обнаруженных изображений людей осуществляется по внешности и может допускать неточности, связанные с разными углами обзора или изменениями освещенности. Однако после обнаружения лица человека на одном из изображений группы перераспределяются, а человеку присваивается уникальный идентификатор. Пример отображения изображений людей, сгруппированных по присвоенным идентификаторам на основе предложенного алгоритма, показан на рисунке 3.

Рисунок 3. – Изображения людей, сгруппированных согласно присвоенным идентификаторам



Для подробного рассмотрения интересующего человека возможен вывод на экран всех полученных изображений, а также информации о месте и времени получения кадра, из которого было извлечено соответствующее изображение (рисунок 4).



**Рисунок 4. – Изображения интересующего человека
с подробной информацией о месте и времени получения изображения**

Заключение. В работе представлен краткий анализ существующих проблем, возникающих при ре-идентификации людей на изображениях систем видеонаблюдения, а также подходов к решению некоторых из них. Уменьшение негативного влияния проблемы смещения домена достигается за счет применения объединенной обучающей выборки. Влияние вариативности освещения, поз, внешности и разрешения камер, окклюзий снижается за счет применения аугментации данных и предложенного алгоритма, учитывающего видимость отдельных фрагментов фигуры человека при формировании дескриптора на основе валидности обнаруженных ключевых точек фигуры. Достигается повышение точности повторной идентификации за счет применяемого метода корректировки гиперпараметров при обучении СНС. Представлена программная реализация системы обнаружения и повторной идентификации человека на изображениях видеонаблюдения путем объединения черт лица и внешнего вида фигуры человека с использованием составного нейронного дескриптора. Для формирования дескриптора лица человека используется обученная сверточная нейронная сеть ResNet50@WebFace600K компании Insightface. Фигура человека описывается составным дескриптором, включающим 1536 признаков для всего изображения, его верхней и нижней частей. Дескриптор формируется с помощью DenseNet-121, обученной с применением рассмотренных подходов, направленных на повышение точности реидентификации. Данный подход позволяет проводить повторную идентификацию человека даже при скрытых лицах за счет анализа признаков всей видимой части фигуры. Он повышает точность в случае перекрытия людей фоновыми объектами и уточняет результат при обнаружении лица человека.

ЛИТЕРАТУРА

1. Deep Learning for Person Re-Identification: A Survey and Outlook / M. Ye, J. Shen, G. Lin et al. // 2021 IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2021. – Vol. 44, iss. 6 – P. 2872–2893. – DOI: [10.1109/TPAMI.2021.3054775](https://doi.org/10.1109/TPAMI.2021.3054775).
2. Brkljac B., Brkljac M. Person detection and re-identification in open-world settings of retail stores and public spaces // ArXiv: 2505.00772. – 2025. – DOI: [10.48550/arXiv.2505.00772](https://doi.org/10.48550/arXiv.2505.00772).
3. Li X., Wu A., Zheng W. Adversarial Open-World Person Re-Identification // 15th European Conference, Munich, Germany, September 8–14, 2018, Proceedings, Part II. – 2018. – P. 287–303. – DOI: [10.1007/978-3-030-01216-8_18](https://doi.org/10.1007/978-3-030-01216-8_18).
4. Illumination Unification for Person Re-Identification / G. Zhang, Z. Luo, Y. Chen et al. // IEEE Transactions on Circuits and Systems for Video Technology. – 2022. – Vol. 32, iss. 10. – P. 6766–6777. – DOI: [10.1109/TCSVT.2022.3169422](https://doi.org/10.1109/TCSVT.2022.3169422).
5. Learning Clothing and Pose Invariant 3D Shape Representation for Long-Term Person Re-Identification / F. Liu, M. Kim, Z. Gu et al. // 2023 IEEE/CVF International Conference on Computer Vision (ICCV). – 2023. – P. 19560–19569. – DOI: [10.1109/ICCV51070.2023.01797](https://doi.org/10.1109/ICCV51070.2023.01797).
6. Person Re-identification in Video Surveillance Systems Using Deep Learning: Analysis of the Existing Methods / H. Chen, S. A. Ihnatsyeva, R. P. Bohush et al. // Automation and Remote Control. – 2023. – Vol. 84, iss. 5. – P. 497–528. – DOI: [10.1134/S0005117923050041](https://doi.org/10.1134/S0005117923050041).
7. Style Normalization and Restitution for Generalizable Person Re-Identification / X. Jin, C. Lan, W. Zeng 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).
8. Generalizable Person Re-Identification by Domain-Invariant Mapping Network / J. Song, Y. Yang, Y. 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).
9. DeVries T., Taylor G. W. Improved regularization of convolutional neural networks with CutOut // ArXiv. – 2017. – DOI: [10.48550/arXiv.1708.04552](https://doi.org/10.48550/arXiv.1708.04552).
10. Dropout: a simple way to prevent neural networks from overfitting / N. Srivastava, G.E. Hinton, A. Krizhevsky et al. // Journal of Machine Learning Research. – 2014. – Vol. 15, iss. 1. – P. 1929–1958. – DOI: [10.5555/2627435.2670313](https://doi.org/10.5555/2627435.2670313).
11. Choice of activation function in convolution neural network for person re-identification in video surveillance systems / H. Chen, S. Ihnatsyeva, R. Bohush et al. // Programming and computer software. – 2022. – № 5. – P. 312–321. – DOI: [10.1134/S0361768822050036](https://doi.org/10.1134/S0361768822050036).

12. Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang et al. // Proceedings of the AAAI Conference on Artificial Intelligence. – 2020. – Vol. 34, iss. 7. – P. 13001–13008. – DOI: [10.1609/aaai.v34i07.7000](https://doi.org/10.1609/aaai.v34i07.7000).
13. CutMix: Regularization Strategy to Train Strong Classifiers With Localizable Features / S. Yun, D. Han, S. Oh Kang et al. // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – P. 6022–6031. – DOI: [10.1109/ICCV.2019.00612](https://doi.org/10.1109/ICCV.2019.00612).
14. Scalable Person Reidentification: A Benchmark / L. Zheng, L. Shen, L. Tian 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).
15. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking / E. Ristani, F. Solera, R. S. Zou et al. // ArXiv. – DOI: [10.48550/arXiv.1609.01775](https://doi.org/10.48550/arXiv.1609.01775).
16. Person Transfer GAN to Bridge Domain Gap for Person Re-identification / L. Wei, S. Zhang, W. Gao et al. // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018. – P. 79–88. – DOI: [10.1109/CVPR.2018.00016](https://doi.org/10.1109/CVPR.2018.00016).
17. Li W., Wang X. Locally Aligned Feature Transforms across Views // 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA. – 2013. – P. 3594–3601. – DOI: [10.1109/CVPR.2013.461](https://doi.org/10.1109/CVPR.2013.461).
18. DeepReID: Deep Filter Pairing Neural Network for Person Re-identification / W. Li, R. Zhao, T. Xiao et al. // 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA. – 2014. – P. 152–159. – DOI: [10.1109/CVPR.2014.27](https://doi.org/10.1109/CVPR.2014.27).
19. Data augmentation and fine tuning of convolutional neural network during training for person re-identification in video surveillance systems / S. Ye, R. Bohush, S. Ihnatsyeva et al. // Optical memory and Neural Network. – 2023. – № 4. – P. 233–246. – DOI: [10.3103/S1060992X23040124](https://doi.org/10.3103/S1060992X23040124).
20. Huang, G. Densely Connected Convolutional Networks / G. Huang, Z. Liu, K. Q. Weinberger et al. // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 2261–2269. – DOI: [10.1109/CVPR.2017.243](https://doi.org/10.1109/CVPR.2017.243).
21. Deep Residual Learning for Image Recognition / K. He, X. Zhang, S. Ren et al. // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – P. 770–778. – DOI: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
22. Miao, J. Pose-Guided Feature Alignment for Occluded Person Re-Identification / J. Miao, Y. Wu, P. Liu et al. // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – P. 542–551. – DOI: [10.1109/ICCV.2019.00063](https://doi.org/10.1109/ICCV.2019.00063).
23. Person re-identification in video surveillance systems by feature replacement of occluded parts of human figures // S. Ye, R. Bohush, S. A. Ihnatsyeva et al. // Pattern Analysis and Applications. – 2025. – Vol. 28. – Art. ID 102. – DOI: [10.1007/s10044-025-01482-1](https://doi.org/10.1007/s10044-025-01482-1).

Поступила 01.10.2025

ALGORITHMIC AND SOFTWARE FOR PERSON RE-IDENTIFICATION BASED ON NEURAL NETWORK FEATURES OF FACE AND FIGURE IMAGES

S. IHNATSYEVA, R. BOHUSH, M. TAMASHEVICH
(Euphrosyne Polotskaya State University of Polotsk)

H. CHEN
(Zhejiang Shuren University, China)

This paper considers the key challenges in implementing open person re-identification systems based on neural network technologies. The influence of hyperparameters during convolutional neural network training on the learning dynamics and accuracy of the person re-identification algorithm is investigated. The experimental selection of neural network hyperparameters for person re-identification consists of two stages. The first stage requires determining the most effective learning rate and image batch size. The second stage involves determining the number of training epochs, taking into account changes in batch size and speed during neural network training. A series of experiments were conducted using ResNet-50 and DenseNet-121 on the PolReID1077, Market-1501, DukeMTMC-ReID, and MSMT17 datasets. Hyperparameters such as batch size, learning rate, and number of neural network training epochs were determined. The experimental results confirmed an increase in re-identification accuracy. In addition, the training time of neural networks using the proposed hyperparameter adjustment method allows for a reduction in training time compared to using the training method on the base model. An algorithm for person re-identification is presented. It uses a descriptor to describe human features based on neural network features of their face and figure. One global and two local neural network descriptors are used to describe a person's facial features. The human figure image is described by a vector of 1536 elements obtained using DenseNet-121. The proposed approach ensures high re-identification accuracy when facial identification is possible and enables re-identification in the case of hidden faces and body parts. Experimental results are presented. A software implementation of a prototype for re-identification in an open world is described.

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