



<http://dx.doi.org/10.35596/1729-7648-2023-21-3-87-95>

Оригинальная статья
Original paper

УДК 004.932

ФОРМИРОВАНИЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ СВЁРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ПРИ РЕИДЕНТИФИКАЦИИ ЛЮДЕЙ ПО ВИДЕОДАНЫМ

С. А. ИГНАТЬЕВА, Р. П. БОГУШ

Полоцкий государственный университет имени Евфросинии Полоцкой (г. Полоцк, Республика Беларусь)

Поступила в редакцию 15.12.2022

© Белорусский государственный университет информатики и радиоэлектроники, 2023
Belarusian State University of Informatics and Radioelectronics, 2023

Аннотация. Для повышения точности работы системы реидентификации людей предлагается комплексный подход при формировании обучающей выборки для свёрточных нейронных сетей, предполагающий использование нового набора изображений, увеличение количества тренировочных примеров за счет существующих баз данных и применение ряда преобразований для повышения их разнообразия. Созданный набор данных PolReID1077 содержит изображения людей, которые были получены во все времена года, что позволит повысить корректность работы систем реидентификации при смене сезонов. Преимуществом PolReID1077 является также использование видеоданных, полученных при внешнем и внутреннем наблюдении в большом количестве различных мест съемки. Поэтому изображения людей в созданном наборе характеризуются вариабельностью фона, яркостных и цветовых характеристик. Объединение созданного набора с существующими CUHK02, CUHK03, Market-1501, DukeMTMC-ReID и MSMT17 позволило получить 109 772 изображения для обучения. Увеличение разнообразия сформированных примеров достигается за счет применения к ним циклического сдвига, исключения цветности и замещения фрагмента уменьшенной копией другого изображения. Представлены результаты исследований по оценке точности реидентификации для свёрточных нейронных сетей ResNet-50 и DenseNet-121 при их тренировке с использованием предложенного подхода для формирования обучающей выборки.

Ключевые слова: повторная идентификация людей, свёрточные нейронные сети, PolReID1077, машинное обучение, аугментация.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Для цитирования. Игнатъева, С. А. Формирование обучающей выборки для свёрточных нейронных сетей при реидентификации людей по видеоданным / С. А. Игнатъева, Р. П. Богуш // Доклады БГУИР. 2023. Т. 21, № 3. С. 87–95. <http://dx.doi.org/10.35596/1729-7648-2023-21-3-87-95>.

TRAINING SAMPLE FORMATION FOR CONVOLUTIONAL NEURAL NETWORKS TO PERSON RE-IDENTIFICATION FROM VIDEO

SVIATLANA A. IHNATSYEVA, RYKHARD P. BOHUSH

Euphrosyne Polotskaya State University of Polotsk (Polotsk, Belarus Republic of Belarus)

Submitted 15.12.2022

Abstract. To improve the person re-identification system accuracy, an integrated approach is proposed in the formation of a training sample for convolutional neural networks, which involves the use of a new image dataset, an increase in the training examples number using existing datasets, and the use of a number of transformations to increase their diversity. The created dataset PolReID1077 contains images of people that were obtained in all

seasons, which will improve the correct operation of re-identification systems when the seasons change. Another PolReID1077 advantage is the video data use obtained from external and internal surveillance in a large number of different filming locations. Therefore, the people images in the created set are characterized by the variability of the background, brightness and color characteristics. Joining the created dataset with the existing CUHK02, CUHK03, Market-1501, DukeMTMC-ReID and MSMT17 sets made it possible to obtain 109 772 images for training. An increase in the variety of generated examples is achieved by applying a cyclic shift to them, eliminating color and replacing a fragment with a reduced copy of another image. The research results on estimating the accuracy of re-identification for the ResNet-50 and DenseNet-121 convolutional neural networks during their training, using the proposed approach to form a training sample, are presented.

Keywords: person re-identification, convolutional neural networks, PolReID1077, machine learning, augmentation.

Conflict of interests. The authors declare no conflict of interests.

For citation. Ihnatsyeva S. A., Bohush R. P. (2023) Training Sample Formation for Convolution Neural Networks to Person Re-Identification from Video. *Doklady BGUIR*. 21 (3), 87–95. <http://dx.doi.org/10.35596/1729-7648-2023-21-3-87-95> (in Russian).

Введение

Широкое распространение систем видеонаблюдения во многих сферах жизнедеятельности людей требует автоматизации процесса анализа видеоданных, среди которых можно выделить актуальность определения присутствия заданного человека на видеоданных в разное время или в другом месте. Такая задача называется повторной идентификацией (реидентификацией) людей и заключается в том, что при поступлении запроса необходимо выделить характерные особенности искомого человека, сравнить их с признаками других людей и определить, когда и где искомый человек попадал в поле зрения множества камер. Наиболее эффективным для этого является применение свёрточных нейронных сетей (СНС), результативность работы которых во многом зависит от качества обучающих данных.

Наборы данных для тренировки нейронной сети должны соответствовать таким требованиям, как большой объем, разнообразие и отсутствие дисбаланса классов в них. Недостаточное количество изображений в выборке может привести к переобучению, т. е. к запоминанию СНС тренировочных примеров и неустойчивости модели к новым данным. Разнообразие набора данных для систем реидентификации определяется количеством видеокамер, различием их характеристик (разрешение, количество кадров в секунду) и вариантами их расположения (место, угол обзора), разными условиями видеонаблюдения (время суток, сезон, освещение). При этом люди на изображениях должны отличаться внешним видом (рост, телосложение, пол, одежда). Увеличение разнообразия повышает робастность извлекаемых признаков и устойчивость обученной СНС к новым изображениям. Дисбаланс классов, при котором одних примеров оказывается существенно больше, чем других, приводит к большему влиянию признаков доминирующего класса при обучении СНС и, соответственно, к потере точности при ее работе.

Для устранения дефицита данных иногда прибегают к объединению существующих наборов или расширению тренировочной выборки за счет имеющихся изображений путем применения к ним каких-либо преобразований [1, 2]. Равномерность распределения классов возможно обеспечить только в процессе формирования набора данных. Сбор, анализ и разметка людей (присвоение уникального номера – идентификатора) – трудоемкая задача, что объясняет ограниченное количество и размеры существующих наборов данных. Наиболее существенные и распространенные из них формировались на основе видео, полученного в течение небольшого временного интервала с 2–15 камер видеонаблюдения при разных условиях видеонаблюдения, и отличаются количеством изображений [3–7]. В связи с этим они обладают различной сложностью для реидентификации и не позволяют получить объективную оценку точности. Наборы данных, имеющиеся в свободном доступе, сформированы в условиях, не схожих с погодными явлениями в регионах, в которых ярко выражена смена сезонов. В случае применения системы реидентификации, для обучения которой использовались подобные изображения, точность работы будет невысокой. Поэтому актуальной задачей является разработка подхода для формирования обучающей выборки на основе комплекса решений, включая создание сбалансированного по временам года набора данных изображений людей, объединение созданного с существующими и расширение разнообразия изображений в общем наборе данных.

Анализ наборов данных для обучения свёрточных нейронных сетей при повторной идентификации человека

Для исследовательских систем повторной идентификации используют наборы данных, содержащие предварительно размеченные изображения людей. Наиболее распространенными являются SUNK02 [3], SUNK03 [4], Market-1501 [5], DukeMTMC-ReID [6], MSMT17 [7]. При формировании набора данных SUNK02 использовалось пять сцен в кампусе университета Гонконга, видеофиксация которых осуществлялась с двух ракурсов. На каждой из пяти сцен получены изображения для 971, 306, 107, 193 и 239 чел., зафиксированных двумя камерами. В зависимости от места, в котором осуществлялось видеонаблюдение, все изображения распределены по каталогам P1, P2, P3, P4 и P5, в каждом из которых есть разделение на Cam1 и Cam2, указывающее на номер камеры. Всего имеется 7264 изображения, предварительно неразмеченных на тестовую и обучающую выборку. Тестирование предложено проводить для изображений 100 случайно выбранных людей, а остальные использовать для обучения.

В кампусе того же университета сформирован набор данных SUNK03. Для 1360 чел. имеется до пяти изображений с каждой из двух камер, которые устанавливались парами для видеонаблюдения в трех различных местах. Сбор данных осуществлялся в течение нескольких месяцев, что дало возможность получить более разнообразные примеры с различными погодными условиями и степенью освещенности. Деление на тестовую и тренировочную выборки осуществлялось по аналогии с SUNK02.

К недостаткам SUNK02 и SUNK03 следует отнести небольшое количество сцен видеонаблюдения и число изображений для каждого человека, случайный характер выбора тестовых изображений в каждом эксперименте. Показатели точности могут отличаться в зависимости от выбранных изображений. Эти недостатки частично были учтены при формировании набора Market-1501, для которого использовались кадры с шести камер видеонаблюдения, расположенных возле супермаркета в университете Циньхуа, включающих 32 668 изображений для 1501 чел. Обучающая выборка содержит 12 936 образцов и 751 идентификатор. Остальные изображения отнесены к тестовой выборке, в том числе 2793 дистрактора. К дистракторам в Market-1501 относятся примеры, на которых детектор ошибочно принял какие-либо объекты за правдоподобные изображения людей. Дополнительно, по запросу, можно получить 500 000 дистракторов, взятых при тех же условиях, что и размеченные изображения. Обнаружение людей на кадрах выполнялось с помощью детектора DPM [8], как и для SUNK03. Качество изображений оценивалось оператором путем сравнения ограничительного прямоугольника, предложенного детектором и отрицанного человеком по метрике IoU [5]. При $\text{IoU} > 50\%$ изображение считалось качественным и включалось в состав набора данных.

DukeMTMC-ReID является подмножеством набора данных Duke Multi-Target Multi-Camera (MTMC), сформированного на основе 85-минутных видео с частотой 60 кадров в секунду. Использовались восемь камер видеонаблюдения, расположенных на территории кампуса университета Дьюк. DukeMTMC разработан для мульти-трекинга и на его основе размечен для задач реидентификации. Для обучения алгоритма повторной идентификации используются 16 522 изображения 702 чел. и 17 661 для 702 других людей для тестирования.

Как Market-1501, так и DukeMTMC-ReID формировались на основе кадров, полученных с наружных камер видеонаблюдения, что ограничивает разнообразие вариантов освещения только естественным светом. Это было учтено при разработке набора данных MSMT17. Для разметки использовались кадры с двенадцати наружных камер видеонаблюдения и трех камер, установленных внутри помещений. Видеофиксация осуществлялась в утреннее, дневное и послеобеденное время в течение четырех дней с различными погодными условиями. В настоящее время набор данных предоставляется с размывом лиц и сопровождается текстовыми файлами, определяющими состав тестовой и обучающей выборок, запросов и валидационных данных, используемых для тестирования во время тренировки СНС. Количество идентифицированных человек составляет 4101, для которых имеется 126 441 изображение. К обучающей выборке относится 32 621 изображение для 1041 чел. При сравнении различных алгоритмов повторной идентификации на разных наборах данных MSMT17 имеет гораздо меньшие показатели точности [9, 10], что указывает на его сложность и низкую способность систем реидентификации повторно верифицировать человека в галерее изображений (тестовой выборке).

Таким образом, разнообразие изображений рассмотренных наборов данных лимитируется небольшим количеством сцен, условий освещения и камер видеонаблюдения. Поэтому СНС запоминает не только специфические особенности людей, но также возможные варианты фона и перекрытия людей элементами ландшафта или интерьера помещений. Предпочтительно, чтобы отличительные признаки, выделенные СНС, являлись характеристикой человека, а не условий видеонаблюдения. Поэтому актуально формирование набора данных для реидентификации с большим количеством различных сцен и вариантов расстановки камер с отличающимися характеристиками.

Методы создания дополнительных изображений на основе существующих

Повысить разнообразие тренировочного набора данных возможно за счет создания дополнительных изображений на основе имеющихся, т. е. путем аугментации. При этом обучающая выборка дополняется изображениями, к которым применялись различные искажения, например, такие, как геометрические или цветовые трансформации: вращение, отражение, размытие, изменение перспективы, цвета, яркости, контраста и др. [1]. Преобразования могут применяться ко всему изображению [11–13], фрагменту [9, 14], пакету изображений [13–16]. В [14] предлагается удалять прямоугольную область изображения путем заполнения его нулевыми или случайными значениями, а в [9] для участка изображения исключается информация о цвете. Некоторые методы аугментации подразумевают комбинирование преобразований по определенному алгоритму [11, 12]. В [11] рассматривается метод RandAugment, при котором одно из четырнадцати преобразований выбирается случайным образом. В [12] для аугментации также используется случайный выбор, но не для одного преобразования, а для цепочки из нескольких искажений, при этом их количество каждый раз может отличаться. Применение аугментации к пакету изображений позволяет смешивать идентификаторы разных людей с помощью линейной интерполяции [13], добавления фрагмента [15] или уменьшенной копии [16] на другое изображение из пакета.

Несмотря на существование большого количества различных методов аугментации, отсутствует достоверный способ определения наиболее эффективного из них [1]. Выбор метода основывается на большом количестве экспериментов с применением различных архитектур нейронных сетей и наборов данных для конкретной задачи.

Формирование обучающей выборки

Формирование набора данных PolReID1077¹ осуществлялось в несколько этапов. На первом выполнялось накопление видео, общая продолжительность которых составляла 14 ч 50 с. Видеопоследовательности получали при содействии студентов и преподавателей Полоцкого государственного университета имени Евфросинии Полоцкой с их согласия только для исследовательских целей. При формировании комплекта изображений для человека обрабатывалось до девяти видеопоследовательностей. Каждый из них был зафиксирован камерами с разных ракурсов в нескольких местах. Использовалось до 10 вариантов расстановки камер для получения изображений каждого человека. Всего для создания набора данных PolReID1077 применялся 1331 способ размещения камер с отличающимися характеристиками (разрешением, количеством кадров в секунду). Видеонаблюдение осуществлялось при разных погодных условиях в течение четырех сезонов (зима – для 30 % задействованных людей, лето – для 29 %, осень и весна – для 41 %). Камеры устанавливались в помещениях с естественным и искусственным освещением, вне помещений – в разное время суток. Наличие у различных людей разных атрибутов в виде сумок, пакетов, рюкзаков, портфелей, продуктовых корзин, шарфов, шапок, очков, папок для бумаг, тетрадей, телефонов, наушников, еды и напитков изменяется с течением времени, что дает возможность изучать и учитывать незначительные изменения внешности. Такой подход к сбору данных отличается от используемых для формирования других наборов и позволяет получить более разнообразные примеры благодаря большому количеству вариантов фоновых помех, уровней освещения, одежды по сезону и т. д. Статистическая информация о распределении данных PolReID1077 приведена в табл. 1.

¹ GitHub. PolReID1077 [Электронный ресурс]. Режим доступа: <https://github.com/SvetlanaIgn/PolReID1077>. Дата доступа: 09.12.2022.

Таблица 1. Статистические данные о составе набора данных PolReID1077
Table 1. Statistic data about the PolReID1077 dataset composition

Характеристика / Characteristic		Количество человек / Persons number	
		Тестовая и обучающая выборки / Test and training samples	Дистрактор / Distractor
Пол	Мужской	705	39
	Женский	302	31
Возраст	18–30 лет	854	57
	Более 30 лет	153	13
Время года	Лето	283	24
	Осень, весна	417	25
	Зима	307	21
Видеосъемка	Внутри помещений	436	31
	Наружная	515	39
	Внутренняя, наружная	156	–
Наличие маски	Да	224	7
	Нет	783	63

На следующем этапе с помощью СНС YOLOv4 выполнялось детектирование людей на кадрах, полученных с интервалом 1–3 с. Каждое изображение маркировалось в формате cXXXXsYY_ZZZZZZ_WW, где XXXX – способ расположения камеры от 0000 до 1331; YY – номер видеопоследовательности, полученной при выбранном способе установки камеры в том же месте, но в другое время; ZZZZZZ – номер кадра с последовательности; WW – количество человек на текущем кадре.

Удаление некорректных обнаружений осуществлялось оператором для случаев: более 50 % фигуры человека было скрыто или находилось вне кадра; детектированы отдельные фрагменты тела; на изображении присутствовали одновременно два и более человек, и было невозможно принять решение, какой идентификатор присвоить; изображения объектов, схожих с силуэтом человека; присутствовали люди, не дававшие согласие на съемку. Формирование комплектов изображений для каждого человека и размещение их в отдельных каталогах также выполнялось оператором. Если изображения были получены на основе кадров с более чем одной камеры, то присваивался уникальный идентификационный номер NNNNN, который дополнял предварительное имя файла, полученное на этапе детектирования, до NNNNN_cXXXXsYY_ZZZZZZ_WW. Каталог с изображениями имеет имя соответствующего идентификатора. Если человек зафиксирован только одной камерой видеонаблюдения, его изображения относятся к дистракторам с меткой 00000.

Для увеличения объема обучающей выборки объединяется несколько наборов данных, схожих по структуре: CUHK02, CUHK03, Market-150, DukeMTMC-ReID, MSMT17 и PolReID1077. Разделение на тестовую и обучающую выборку для Market-1501, MSMT17, DukeMTMC-ReID и PolReID1077 осуществляется в соответствии с рекомендациями разработчиков. CUHK02 и CUHK03 полностью отнесены к обучающей выборке и не используются для тестирования. Суммарно для тренировки СНС использовалось 109 772 изображения для 6381 чел.

Для повышения разнообразия созданного набора данных предлагается метод аугментации, который включает циклический сдвиг пикселей по вертикали и горизонтали, исключение цветности и замещение фрагмента уменьшенным изображением из пакета. Первая группа преобразований, к которой относятся циклический сдвиг пикселей и исключение цветности, выполняется в процессе формирования пакета. Для изображения с помощью генератора псевдослучайных чисел (ГПСЧ) определяется необходимость применения каждого из преобразований. После формирования пакета перед его поступлением на вход нейронной сети аналогичным образом принимается решение о замещении фрагментов уменьшенными копиями изображений из пакета.

В результате применения аугментации обучающая выборка на каждой эпохе обучения формируется с различными преобразованиями. На рис. 1, *a* показан пример сформированного пакета с аугментацией, при которой для случайно выбранных изображений выполняются циклический сдвиг и исключение цветности. Рис. 1, *b* демонстрирует пример пакета, в котором в изображение включена уменьшенная копия другого.

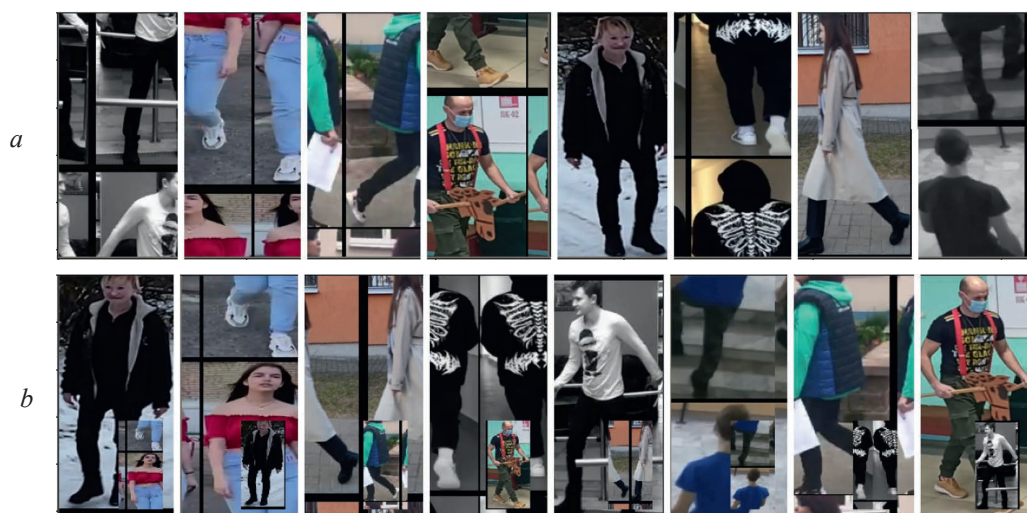


Рис. 1. Примеры изображений людей из набора данных PolReID1077: *a* – пакет изображений формируется с применением циклического сдвига и исключением цветности;

b – уменьшенная копия добавляется на другое изображение из пакета

Fig. 1. Sample of persons images from PolReID1077 dataset: *a* – images batch is formed using a cyclic shift and color exclusion; *b* – a reduced copy is added to another image from the batch

Такой метод аугментации исключает возможность запоминания тренировочных данных, повышает устойчивость СНС к окклюзиям, освещенности, разрешению изображения, зависимости от местоположения признаков.

Методика проведения эксперимента

Для проведения экспериментов в качестве базового применялся алгоритм повторной идентификации², реализованный на фреймворке Pytorch. В качестве экстракторов признаков рассматривались СНС ResNet-50 и DenseNet-121. Точность реидентификации оценивалась по трем метрикам Rank1, mAP и mINP. Исследования проводились на персональном компьютере с характеристиками: Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb. С целью повышения точности реидентификации использовался подход двухэтапного обучения, основная идея которого заключалась в том, что на первом этапе для обучения применялись аугментированные данные, на втором выполнялась точная настройка весовых коэффициентов на исходных изображениях. Для предобучения формировались пакеты размером $B = 16$. Часть пакетов содержала изображения (\tilde{X}, \tilde{Y}) с замещенными фрагментами, остальные состояли из изображений (X_i, Y_j) , которые могут быть с циклическим сдвигом или полутоновыми. Изображения и пакеты, к которым применялось преобразование, определялись с помощью ГПСЧ, генерирующего псевдослучайное число $r \in 0 \dots 1$. Пороговое значение вычислялось по формуле

$$t = \frac{p}{100}, \quad (1)$$

где p – гиперпараметр, определяющий, к какой части изображений будет применяться преобразование; если $r < t$, то изображение (пакет) видоизменяется.

Значения функции потерь L_{batch} для обрабатываемых пакетов накапливаются в течение эпохи, после чего вычисляется средняя величина ошибки

$$L_{epoch} = \frac{\sum L_{batch} B}{size_{dataset}}, \quad (2)$$

где $size_{dataset}$ – размер обучающей выборки.

² Github. Pytorch ReID. Available: https://github.com/layumi/Person_reID_baseline_pytorch (Accessed 9 Desember 2022).

Если пакет содержит уменьшенные копии других изображений, то

$$L_{batch} = E(\tilde{X}^{out}, Y_i)\lambda + E(\tilde{X}^{out}, Y_j)(1 - \lambda), \quad (3)$$

где $\lambda = 1 - \frac{r_w r_h}{WH}$ – коэффициент, зависящий от соотношения размеров уменьшенной копии и исходного изображения; r_w, r_h – ширина и высота уменьшенной копии; W, H – ширина и высота изображения; $E()$ – кросс-энтропийная функция потерь.

Если при формировании пакета добавление уменьшенных копий не применялось, то

$$L_{batch} = E(X_i^{out}, y_i). \quad (4)$$

Весовые коэффициенты w корректируются со скоростью обучения $lr = 0,07$

$$w = w - lr \frac{\partial L_{epoch}}{\partial w}. \quad (5)$$

После 40-й эпохи происходит уменьшение скорости обучения в 0,1 раза, а предварительный этап длится 45 эпох. На втором этапе выполняется точная настройка в течение 35 эпох, для которой формирование пакетов B изображений (X_i, Y_j) реализуется без аугментации. Накопление L_{batch} и вычисление значения функции потерь для всей эпохи L_{epoch} выполняются аналогично первому этапу, как и корректировка весовых коэффициентов. Такой подход позволяет уменьшить значение функции потерь и повысить точность повторной идентификации.

Каждой из двух СНС соответствуют три эксперимента.

1. Данные для обучения и тестирования принадлежат одному домену (набору данных).
2. Для обучения использовался объединенный набор данных, для тестирования – соответствующие выборки Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077.
3. Обучение осуществлялось на объединенном наборе данных с применением аугментации. Точность оценивалась на тестовых выборках наборов данных Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077.

Результаты экспериментов

В табл. 2 представлены результаты тестирования СНС на наборах данных Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077 с применением различных подходов к обучению. В качестве эталонных значений принимали показатели Rank1, mAP и mINP, полученные для СНС, обучение и тестирование которых выполняли на данных, принадлежащих одному домену. Увеличение количества изображений позволяет повысить точность повторной идентификации по трем метрикам оценки точности для каждого из исследуемых наборов данных.

Таблица 2. Сравнение точности работы алгоритма повторной идентификации людей при использовании для тренировки свёрточной нейронной сети различных способов формирования обучающей выборки
Table 2. Comparison of persons re-identification accuracy when using different methods of forming a sample to train the convolutional neural network

СНС / CNN	Обучающая выборка / Train sample	Подход к обучению / Training approach	Метрика / Metric	Тестовая выборка / Test sample			
				Market-1501	DukeMTMC-ReID	MSMT17	PolReID1077
ResNet-50	Оригинальная*	Базовый алгоритм	Rank1	83,19	72,85	49,54	88,94
			mAP	61,08	52,53	24,83	65,58
			mINP	23,95	14,53	2,51	20,86
	Объединенная	Базовый алгоритм	Rank1	91,75	82,72	69,49	95,56
			mAP	77,64	69,03	42,57	81,27
			mINP	46,32	31,24	07,05	40,29
Объединенная	С применением аугментации	Rank1	92,28	85,05	73,24	96,25	
		mAP	81,34	72,21	47,79	84,00	
		mINP	52,78	35,38	9,67	45,55	

Окончание табл. 2
Ending of Tab. 2

СНС / CNN	Обучающая выборка / Train sample	Подход к обучению / Training approach	Метрика / Metric	Тестовая выборка / Test sample			
				Market-1501	DukeMTMC-ReID	MSMT17	PolReID1077
DenseNet-121	Оригинальная*	Базовый алгоритм	Rank1	83,76	72,80	54,98	90,33
			mAP	62,48	53,16	26,67	68,39
			mINP	26,83	15,88	2,61	23,43
	Объединенная	Базовый алгоритм	Rank1	92,13	83,84	73,51	96,25
			mAP	79,48	70,30	46,76	82,90
			mINP	48,92	31,66	8,18	43,14
	С применением аугментации	Rank1	92,75	85,50	75,10	96,94	
		mAP	82,03	72,75	50,13	85,25	
		mINP	53,40	35,26	10,24	47,53	

* Под оригинальной обучающей выборкой предполагается использование для обучения того же набора данных, что и для тестирования.

Применение для тренировки СНС обучающей выборки, расширенной за счет объединения нескольких наборов данных и аугментации, позволило значительно повысить точность повторной идентификации для СНС ResNet-50 в метрике Rank1 на 8–48 %, в метрике mAP – на 27–91 %, в mINP – на 118–285 %. Для DenseNet-121 достигнуто увеличение точности на 7–37 % в метрике Rank1, на 25–88 % – в mAP, на 99–290 % – в mINP при тестировании на разных наборах данных.

Выводы

1. Предложен комплексный подход формирования обучающей выборки для увеличения точности повторной идентификации людей в распределенных системах видеонаблюдения. Для этого создан новый набор данных PolReID1077, который отличается от существующих тем, что содержит изображения людей, полученные в различные времена года и во множестве мест съемки. Объединение с созданной выборкой наиболее известных наборов данных, применяемых для обучения сверточной нейронной сети при реидентификации, позволило увеличить количество изображений на 236,5 % относительно наибольшего из применяемых MSMT17. Использование предложенного алгоритма аугментации данных повышает устойчивость сверточной нейронной сети к перекрытиям, степени освещенности, разрешению изображений и вариациям поз людей.

2. Комплексный подход к формированию обучающей выборки для обучения сверточной нейронной сети позволил повысить точность реидентификации в метриках Rank1, mAP и mINP для разных архитектур сверточной нейронной сети и наборов данных.

References

- Shorten C., Taghi M. K. (2019) A Survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*. (6), 1–48. DOI: 10.1186/s40537-019-0197-0.
- Chen H., Ihnatsyeva S., Bohush R., Ablameyko S. (2022) Choice of Activation Function in Convolution Neural Network in Video Surveillance Systems. *Programming and Computer Software*. (5), 312–321. DOI: 10.1134/S0361768822050036.
- Wei L., 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.
- 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.
- 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*. 1116–1124. DOI: 10.1109/ICCV.2015.133.
- Ristani E., Solera F., Zou R. S., Cucchiara R., Tomasi C. (2016) Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. *ECCV Workshops*. DOI: 10.1007/978-3-319-48881-3_2.
- Wei L., Zhang S., Gao W., Tian Q. (2017) 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.

8. Felzenszwalb P., Girshick R., McAllester D., Ramanan D. (2010) Object Detection with Discriminatively Trained Part-Based Models. *PAMI*. DOI: 10.1109/TPAMI.2009.167.
9. Gong Y., Zeng Z. (2021) An Effective Data Augmentation for Person Re-Identification. *ArXiv, abs/2101.08533*. DOI: 10.48550/arXiv.2101.08533.
10. Fu D., Chen D., Bao J., Yang H., Yuan L., Zhang L., Li H., Chen D. (2021) Unsupervised Pre-Training for Person Re-Identification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14745–14754. DOI: 10.1109/CVPR46437.2021.01451.
11. Cubuk E. D., Zoph B., Shlens J., Le Q. V. (2020) Randaugment: Practical Automated Data Augmentation with a Reduced Search Space. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 3008–3017. DOI: 10.1109/CVPRW50498.2020.00359.
12. Hendrycks D., Mu N., Cubuk E. D., Zoph B., Gilmer J., Lakshminarayanan B. (2019) AugMix: a Simple Data Processing Method to Improve Robustness and Uncertainty. *ArXiv, abs/1912.02781*. DOI: 10.48550/arXiv.1912.02781.
13. Zhang H., Cissé M., Dauphin Y., Lopez-Paz D. (2017) Mixup: Beyond Empirical Risk Minimization. *ArXiv, abs/1710.09412*. DOI: 10.48550/arXiv.1710.09412.
14. Zhong Z., Zheng L., Kang G., Li S., Yang Y. (2017) Random Erasing Data Augmentation. *AAAI Conference on Artificial Intelligence*. DOI: 10.1609/AAAI.V34I07.7000.
15. Yun S., Han D., Oh S., Chun S., Choe J., Yoo Y. J. (2019) CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. *2019 IEEE/CVF International Conference on Computer Vision*. 6022–6031. DOI: 10.1109/ICCV.2019.00612.
16. Xie T., Cheng X., Wang X., Liu M., Deng J., Zhou T., Liu M. (2021) Cut-Thumbnail: a Novel Data Augmentation for Convolutional Neural Network. *Proceedings of the 29th ACM International Conference on Multimedia*. DOI: 10.1145/3474085.3475302.

Вклад авторов / Authors' contribution

Все авторы внесли равный вклад в написание статьи / The authors contributed equally to the writing of the article.

Сведения об авторах

Игнатьева С. А., аспирант кафедры вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой

Богуш Р. П., д. т. н., доцент, заведующий кафедрой вычислительных систем и сетей Полоцкого государственного университета имени Евфросинии Полоцкой

Адрес для корреспонденции

211440, Республика Беларусь,
г. Новополоцк, ул. Блохина, 29
Полоцкий государственный университет
имени Евфросинии Полоцкой
Тел.: +375 214 42-30-31
E-mail: s.ignatieva@pdu.by
Игнатьева Светлана Александровна

Information about the authors

Ihnatsyeva S. A., Postgraduate at the Department of Computing Systems and Networks of the Euphrosyne Polotskaya State University of Polotsk

Bohush R. P., Dr. of Sci. (Tech.), Associate Professor, Head of the Department of Computing Systems and Networks of the Euphrosyne Polotskaya State University of Polotsk

Address for correspondence

211440, Republic of Belarus,
Novopolotsk, Blokhina St., 29
Euphrosyne Polotskaya
State University of Polotsk
Tel.: +375 214 42-30-31
E-mail: s.ignatieva@pdu.by
Ihnatsyeva Sviatlana Aleksandrovna