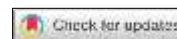


ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

INTELLIGENT SYSTEMS



УДК 004.932

<https://doi.org/10.37661/1816-0301-2023-20-1-40-54>Оригинальная статья
Original Paper

Увеличение точности реидентификации людей на основе двухэтапного обучения сверточных нейронных сетей и аугментации

С. А. Игнатьева[✉], Р. П. Богуш*Полоцкий государственный университет имени Евфросинии Полоцкой,
ул. Блохина, 29, Новополоцк, 211440, Беларусь*[✉]E-mail: s.ignatieva@pdu.by

Аннотация

Цели. Основной целью исследования является повышение точности повторной (ре-) идентификации людей в распределенных системах видеонаблюдения.

Методы. Применяются методы машинного обучения.

Результаты. Представлена технология двухэтапного обучения сверточных нейронных сетей (СНС), отличающаяся использованием аугментации изображений для предварительного этапа и точной настройки весовых коэффициентов на основе исходного набора изображений. На первом этапе обучение осуществляется на аугментированных данных, на втором этапе выполняется точная настройка СНС на исходных изображениях, что способствует повышению эффективности реидентификации за счет уменьшения потерь при обучении. Применение на двух этапах разных данных не позволяет СНС запоминать тренировочные примеры, тем самым предотвращая переобучение.

Предложенный метод расширения набора данных для обучения отличается тем, что совмещает циклический сдвиг пикселей изображения, исключение цветности и замещение одного фрагмента изображения уменьшенной копией другого из пакета, подаваемого на вход СНС. При использовании этого метода аугментации можно увеличить разнообразие обучающих данных, что повысит робастность СНС ко многим факторам: перекрытию людей, изменению освещенности, уменьшению разрешения изображения, зависимости от местоположения отличительных особенностей объекта интереса.

Заключение. Применение технологии двухэтапного обучения и предложенного метода аугментации данных позволяет повысить точность повторной идентификации людей для разных СНС и наборов данных в метриках: Rank1 на 4–21 %, mAP на 10–31 %, mINP на 39–60 %.

Ключевые слова: реидентификация людей, сверточные нейронные сети, предварительное обучение, точная настройка, расширение обучающей выборки

Для цитирования. Игнатьева, С. А. Увеличение точности реидентификации людей на основе двухэтапного обучения сверточных нейронных сетей и аугментации / С. А. Игнатьева, Р. П. Богуш // Информатика. – 2023. – Т. 20, № 1. – С. 40–54. <https://doi.org/10.37661/1816-0301-2023-20-1-40-54>

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию | Received 05.10.2022
Подписана в печать | Accepted 25.11.2022
Опубликована | Published 29.03.2023

Improving person re-identification based on two-stage training of convolutional neural networks and augmentation

Sviatlana A. Ihnatsyeva[✉], Rykhard P. Bohush

*Euphrosyne Polotskaya State University of Polotsk,
st. Blokhina, 29, Novopolotsk, 211440, Belarus*

[✉]*E-mail: s.ignatieva@pdu.by*

Abstract

Objectives. The main goal is to improve person re-identification accuracy in distributed video surveillance systems.

Methods. Machine learning methods are applied.

Result. A technology for two-stage training of convolutional neural networks (CNN) is presented, characterized by the use of image augmentation for the preliminary stage and fine tuning of weight coefficients based on the original images set for training. At the first stage, training is carried out on augmented data, at the second stage, fine tuning of the CNN is performed on the original images, which allows minimizing the losses and increasing model efficiency. The use of different data at different training stages does not allow the CNN to remember training examples, thereby preventing overfitting.

Proposed method as expanding the training sample differs as it combines an image pixels cyclic shift, color exclusion and fragment replacement with a reduced copy of another image. This augmentation method allows to get a wide variety of training data, which increases the CNN robustness to occlusions, illumination, low image resolution, dependence on the location of features.

Conclusion. The use of two-stage learning technology and the proposed data augmentation method made it possible to increase the person re-identification accuracy for different CNNs and datasets: in the Rank1 metric by 4–21 %; in the mAP by 10–31 %; in the mINP by 39–60 %.

Keywords: Person re-identification, convolutional neural network, pre-train, fine tuning, augmentation

For citation. Ihnatsyeva S. A., Bohush R. P. *Improving person re-identification based on two-stage training of convolutional neural networks and augmentation*. Informatika [Informatics], 2023, vol. 20, no. 1, pp. 40–54 (In Russ.). <https://doi.org/10.37661/1816-0301-2023-20-1-40-54>

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

Введение. СНС являются наиболее эффективным инструментом извлечения признаков изображений людей для реидентификации в многокамерных системах видеонаблюдения. Известно, что в значительной мере на эффективность работы СНС оказывают влияние количество изображений в обучающей выборке и их разнообразие. Недостаток тренировочных данных может спровоцировать переобучение СНС, запоминание исходных данных и неспособность к обобщению признаков в целом. Решением этой проблемы может служить концепция переноса обучения (transfer learning). В таком случае СНС обучается на наборе данных достаточно большого размера, например ImageNet (3,2 млн аннотированных изображений) [1] или LUPerson (около 4 млн изображений более 200 тыс. человек) [2]. После обучения возможны два сценария (URL: https://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html): использование обученной модели в качестве сети с фиксированными связями (fixed feature extractor) и выполнение точной настройки (fine tuning). В первом случае обучение СНС на исследуемом наборе данных выполняется только для полносвязного слоя, а весовые коэффициенты скрытых слоев остаются без изменений. При точной настройке СНС скрытые слои обучаются с меньшей скоростью, чем последний слой, который является классификационным. Таким образом, СНС обучается в два этапа. На первом этапе выполняется предварительное обучение на большом наборе данных, а на втором этапе веса корректируются на обучающей выборке для решения конкретной задачи, например обнаружения и классификации объектов разных классов, повторной идентификации человека [3].

Эффективность предобученных моделей СНС подтверждена результатами исследований, опубликованными в работах [2–4], и обусловлена тем, что на этапе предварительного обучения выделяются признаки, несущие основную информацию об изображениях: вертикальные и горизонтальные линии, цвет, текстура и форма объектов и др. Использование предварительно обученных нейронных сетей широко распространено не только для решения задач классификации и распознавания объектов на изображениях, но и для сегментации изображений, поиска ключевых точек и обработки текста (URL: <https://pytorch.org/vision/stable/models.html>). Применение таких сетей ограничивается наличием предварительно обученных моделей для конечного числа архитектур СНС и высоким потреблением вычислительных ресурсов и времени.

Другим решением проблемы переобучения является регуляризация. В машинном обучении под регуляризацией понимают добавление ограничений к архитектуре нейронной сети, функции потерь или к обучающим наборам данных [5]. Примерами регуляризации являются: прореживание нейронных связей СНС [6]; использование разных функций активации [7]; L1- и L2-регуляризации (URL: <https://www.analyticssteps.com/blogs/l2-and-l1-regularization-machine-learning>); аугментация данных, при которой часть изображения может быть удалена или замещена другой информацией [5, 8, 9]. Аугментация применяется для увеличения разнообразия обучающей выборки на основе имеющихся данных [9] за счет преобразований изображений. При этом могут использоваться такие преобразования, как изменение яркости и контраста, зеркальное отражение, поворот, размытие и др. Количество изображений при аугментации не увеличивается, а разнообразие достигается за счет того, что на различных эпохах обучения к изображениям применяются разные преобразования. Расширение обучающей выборки позволяет улучшить обобщающую способность СНС и увеличить точность работы, в том числе при различных мешающих факторах, таких как высокая вариация освещения, низкое разрешение, перекрытие объектов и др.

Анализ существующих методов аугментации. При увеличении обучающей выборки преобразование выполняется для одного изображения [8] или для пакета изображений [9–11], используемых при обучении СНС. Изображения группируются в пакеты для уменьшения времени обучения за счет распараллеливания вычислений. Размер пакета является гиперпараметром и, как правило, изменяется в диапазоне от 16 до 64. Следует отметить, что уменьшение размера пакета может приводить к взрывным градиентам и неспособности СНС к обобщению, а увеличение – к снижению точности работы СНС.

В работе [8] для аугментации предлагается метод *random erasing*, при котором изображения выбираются для преобразований из исходного набора данных с применением генератора псевдослучайных чисел (ГПСЧ). На основе ГПСЧ также определяются размер и координаты прямоугольного фрагмента изображения, пиксели которого заполняются нулевыми или случайными значениями (рис. 1, *a*). Поэтому при загрузке пакета изображений на разных эпохах обучения одно и то же изображение может быть представлено как в исходном виде, так и с разными измененными фрагментами, что позволяет повысить устойчивость СНС к окклюзиям, однако при этом часть информации теряется.

Метод регуляризации с использованием аугментации *CutMix* основан на применении преобразования в пакете изображений, которое предполагает замещение области фрагмента одного изображения областью такого же размера другого изображения из этого же пакета [9] (рис. 1, *b*). Применение *CutMix* позволяет СНС повысить устойчивость к перекрытию и при этом сохранить ту часть информации, которая теряется при использовании метода *random erasing*. В алгоритме *Mixed Single Thumbnail (MST)* [10] фрагмент изображения замещается на уменьшенную копию фиксированного размера другого изображения из пакета (рис. 1, *c*). Это позволяет одновременно изучать два изображения при извлечении признаков, повышает устойчивость к окклюзиям и низкому разрешению изображений. Одновременно два изображения из пакета обрабатываются и при использовании аугментации *mixup* [11], однако их объединение выполняется с помощью линейной интерполяции (рис. 1, *d*). Такой подход позволяет улучшить обобщающую способность обученной СНС и снижает чувствительность к запоминанию ошибочных меток (идентификаторов людей), которые могут встречаться в существующих наборах данных.



a)



b)



c)



d)

Рис. 1. Примеры использования методов аугментации данных: a) random erasing; b) CutMix; c) MST; d) mixup
Fig. 1. Examples of using data augmentation methods: a) random erasing; b) CutMix; c) MST; d) mixup

Результаты исследований, представленные в работах [12, 13], показывают, что при обучении для задачи классификации нейронная сеть больше внимания уделяет текстуре объекта, а не его форме. Так, в [12] отмечено, что если образ кота заполнить текстурой кожи слона, то СНС распознает класс обнаруженного объекта как «слон», тогда как человек будет считать, что это кот, т. е. для человека решающее значение имеет именно форма объекта, а не его текстура.

Возможны ситуации, когда люди носят одежду одного фасона и разного цвета, причем даже у одного человека цвет одежды при изменении освещения может различаться. В таком случае при реидентификации важное значение будет иметь силуэт человека. Поэтому для увеличения значимости формы объекта на предсказание СНС в статье [13] предлагается исключать цветность на прямоугольных фрагментах изображения.

В работе [14] предлагается метод аугментации, состоящий из нескольких этапов. Вначале СНС обучается на исходных данных до того момента, когда значения функции потерь перестают уменьшаться. После обучения на изображениях определяются области, которые являются наиболее важными для принятия решения нейронной сетью. Для этого методом скользящего окна исключаются разные участки изображения и оценивается влияние данного шага на ре-

зультат. При обнаружении фрагмента изображения, без которого сеть не может верно определить идентификационную метку человека, область исключается и изображение добавляется к обучающей выборке. После расширения обучающей выборки изображениями с удаленными фрагментами продолжается обучение СНС. Применение такого метода аугментации дает возможность обучить сеть нахождению дополнительных отличий людей, позволяющих повысить точность повторной идентификации.

Методы random erasing и CutMix повышают устойчивость лишь к окклюзиям, а MST – к окклюзиям и разрешению изображения. Метод, предложенный в работе [14], требует дополнительных затрат на поиск изображений, которые будут использоваться для аугментации.

Для исключения указанных недостатков существующих подходов и повышения эффективности повторной идентификации людей в системах распределенного видеонаблюдения предлагается технология двухэтапного обучения с помощью нового метода аугментации данных.

Технология двухэтапного обучения с аугментацией данных. Предлагается использовать подход, включающий технологию двухэтапного обучения СНС и новый метод аугментации данных. При этом на первом этапе выполняется предварительное обучение на аугментированных данных, а на втором этапе применяется точная настройка СНС, при которой весовые коэффициенты корректируются на исходных изображениях с меньшей скоростью обучения. Это позволяет извлекать наиболее надежные отличительные признаки. Для увеличения разнообразия обучающей выборки на основе имеющихся данных предлагается использовать циклический сдвиг пикселей по вертикали и горизонтали, исключение цветности и замещение фрагмента другим уменьшенным изображением. Циклический сдвиг и исключение цветности применяются к отдельным изображениям, а замещение фрагмента осуществляется уменьшенной копией другого из пакета, подаваемого на вход СНС.

Предварительное обучение модели и точная настройка. Для снижения значения функции потерь при обучении, высокие уровни которой вызваны неправдоподобностью обучающих данных по отношению к тестовым данным, и повышения точности повторной идентификации людей предлагается применять двухэтапное обучение. При этом используются следующие особенности: скорость тренировки классификационного слоя выше скорости всех остальных слоев СНС, на первом этапе осуществляется предварительное обучение с применением аугментации данных, на втором этапе тренировка СНС продолжается только на исходном наборе данных.

Циклический сдвиг изображения. В большинстве случаев голова человека находится в верхней части изображения, а ноги – в нижней. В процессе обучения СНС запоминает эти особенности. При обнаружении человека детектором часть его изображения может не попасть в ограничительную рамку или быть скрыта другими объектами. Если человек наклонился, присел, поднял руки вверх, то нижняя или верхняя часть его тела оказывается скрыта, поэтому выделение признаков СНС выполняется не совсем корректно. Применение циклического сдвига позволяет получить изображение человека, на котором отдельные части тела могут оказаться не на своих местах и отдельно от самого человека. Например, голова будет находиться в нижней части изображения, ступни – выше плеч, а правая рука – слева. Представляется, что такой подход обеспечит повышение устойчивости СНС к местоположению отличительных характеристик фигуры.

Циклический сдвиг применяется к некоторой части изображений (p) из всего набора исходных данных до начала обучения как гиперпараметр наряду с размером пакета B , количеством эпох $K_{\text{эпох}}$ и скоростью обучения lr . Для реализации выбора изображений (листинг), к которым применяется данное преобразование, используется ГПСЧ, с помощью которого определяется число r , принимающее значение от нуля до единицы. Пороговое значение

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

служит для определения необходимости применения циклического сдвига к изображению. Если $r < t$, то входное изображение видоизменяется.

Листинг. Алгоритм выбора изображений из обучающей выборки, для которых выполняется преобразование

```

1  Input:  $I_{in}$  – входное изображение
2      p – процент изображений, к которым применяется преобразование
3      Transform – преобразование
4  Output:  $I_{out}$  – выходное изображение
5  Initialization:  $r \leftarrow \text{Rand}(0,1)$ ;
6       $t = p/100$ 
7  if  $r > t$  then
8       $I_{out} \leftarrow I_{in}$ ;
9      return  $I_{out}$ .
10 else
11      $I_{out} \leftarrow \text{Transform}(I_{in})$ ;
12     return  $I_{out}$ .
13 end
    
```

На рис. 2 показаны результаты циклического сдвига по вертикали и горизонтали на некоторое количество пикселей изображений.



Рис. 2. Примеры циклического сдвига пикселей изображений по вертикали и горизонтали

Fig. 2. Examples of image pixel's cyclic shift vertically and horizontally

Исключение цветности. Цвет одежды является важной характерной особенностью, играющей значительную роль при реидентификации людей. Однако цветопередача различных видеокамер и освещенность сцены видеонаблюдения могут значительно исказить цветовые характеристики одного и того же объекта: на рис. 3, *a* при разном освещении платье женщины воспроизводится серым цветом вместо белого; на рис. 3, *b* с одной камеры майка фиксируется голубой, с другой – бирюзовой; пример на рис. 3, *c* демонстрирует, как коричневые шорты меняют цвет ближе к черному. В этом случае наиболее важными представляются такие признаки, как силуэт человека или форма одежды. Поэтому исключение цветности может повысить устойчивость к уровням освещения за счет увеличения влияния других признаков изображения человека.



Рис. 3. Примеры изображений с видеокамер при разном освещении

Fig. 3. Color distortion examples in different lighting conditions

Замещение фрагмента одного изображения уменьшенной копией другого из пакета.

Добавление копии одного изображения на другое выполняется методом, похожим на метод из работы [10]. Его отличие заключается в том, что уменьшенные копии имеют нефиксированный размер: изменяют значения в диапазоне от $0,125S$ до $0,25S$, где S – площадь исходного изображения.

На входе СНС изображение x_i имеет метку (идентификатор) y_i , где i – индекс изображения в пакете. На каждом изображении (x_i, y_i) фрагмент замещается на $(x_j^{\text{mini}}, y_j^{\text{mini}})$. Индекс j – порядковый номер изображения в пакете, которое будет уменьшено; x_j^{mini} – уменьшенная копия x_j ; y_j^{mini} – соответствующий идентификатор. Для определения индекса j для x_j^{mini} , замещающего фрагмент на i -м изображении, с помощью ГПСЧ формируется массив J размером B , заполненный числами от 1 до B , где B – количество изображений в пакете без повторений. Полученное в результате преобразования изображение можно описать следующим образом:

$$\tilde{x} = M \odot x_i + \Phi(x_j^{\text{mini}}), \quad \tilde{y} = \lambda y_i + (1 - \lambda) y_j^{\text{mini}}, \quad (2)$$

где (\tilde{x}, \tilde{y}) – аугментированное изображение; $M \in \{0,1\}^{W \times H}$ – бинарная маска, определяющая местоположение фрагмента, который будет заменен на уменьшенное изображение из пакета; W и H – ширина и высота x_i ; \odot – поэлементное умножение; $\Phi(\cdot)$ – операция заполнения изображения.

Для бинарной маски необходимо сформировать координаты ограничительной рамки вырезанного фрагмента $E = (r_x, r_y, r_w, r_h)$, где r_x, r_y – координаты левого нижнего угла; r_w и r_h – высота и ширина удаленного участка, которые определяются с помощью ГПСЧ таким образом, что r_h составляет $(0,25 \dots 0,5 H)$, где H – высота исходного изображения. Ширина определяется как $r_w = \left\lfloor \frac{r_h}{\eta} \right\rfloor$, где $\eta = \left\lfloor \frac{H}{W} \right\rfloor$ – соотношение сторон исходного изображения.

Область размером E на x_i удаляется и заполняется уменьшенной копией x_j^{mini} такого же размера, что и E . Для $\Phi(\cdot)$ формируется бинарная маска $\tilde{M} = 1 - M$, заполненная единицами, что позволяет разместить в удаленной области x_i уменьшенную копию x_j^{mini} . При таком методе аугментации на каждом изображении присутствуют два человека с соответствующими метками y_i и y_j^{mini} . Метка \tilde{y} будет зависеть от соотношения площади уменьшенного изображения к исходному, определяемого коэффициентом λ :

$$\lambda = 1 - \frac{r_w \cdot r_h}{W \cdot H}. \quad (3)$$

Рассмотрим пример применения предлагаемого метода аугментации для пакета из восьми изображений. Изображения на рис. 4, *a–d, f* и *h* содержат циклический сдвиг; на изображении на рис. 4, *b* отсутствует информация о цвете. Для всех изображений выполнена замена фрагмента на уменьшенное изображение из пакета. Уменьшенные изображения содержат циклический сдвиг в примерах на рис. 4, *a, d–h*.



Рис. 4. Примеры аугментации данных с циклическим сдвигом пикселей, добавлением изображений в градациях серого и замещением фрагмента изображения уменьшенной копией

Fig. 4. Examples of data augmentation with cycle shift pixel, adding grayscale images and replacing image fragment with a reduced copy

Представленные на рис. 4 изображения в значительной мере отличаются от реальных, полученных при видеонаблюдении. Предполагается, что с их помощью можно обучить СНС обнаруживать признаки независимо от их расположения, целостности частей тела человека, попавших в кадр, цвета одежды или разрешения изображения и таким образом повысить чувствительность СНС к форме (силуэту) объекта, увеличить устойчивость к окклюзиям, уменьшить зависимость от уровня освещенности.

Оценка точности работы алгоритма повторной идентификации людей. Для оценки результатов повторной идентификации использовались такие метрики, как mAP, Rank1 и mINP [15].

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

$$mAP = \frac{1}{Q} \sum_{i=1}^Q AP_i, \quad (4)$$

где Q – общее число запросов; $AP = \frac{\sum \text{precision}}{I}$ – средняя точность для каждого i -го запроса;

I – число изображений в тестовой выборке; $\text{precision} = \frac{TP}{TP + FP}$ – точность запроса; TP – количество верноположительных предсказаний запроса; FP – количество ложноположительных предсказаний запроса.

Группа метрик RankN характеризует качество ранжирования и показывает проценты от общего числа запросов, для которых верный выданный результат был среди первых N полученных. Соответственно, метрика Rank1 показывает число процентов запросов, для которых идентификатор первого изображения-кандидата совпадает с идентификатором запроса. Если $N = 5$, то Rank5 показывает число процентов запросов, для которых среди первых пяти выданных изображений-кандидатов было определено верное решение. Для вычисления RankN определяется отношение суммы числа запросов, для которых верное решение было найдено среди первых выданных результатов, к общему числу запросов Q :

$$\text{RankN} = \frac{\sum K_{i,N}}{Q}, \quad (5)$$

где i – номер запроса; $K_{i,N}$ – i -й запрос, для которого верное решение было получено в ряду первых N выданных результатов.

В системах повторной идентификации приоритетно, чтобы верные предсказания находились в начале ранжированного списка и имели как можно меньше ложных предсказаний. Метрики RankN и mAP не отражают сложность поиска правильно идентифицированных изображений людей для поступающего запроса, и при одинаковых показателях Rank значения точности AP могут отличаться. Для учета поиска наиболее сложных правильных предсказаний используется метрика mINP (mean Inverse Negative Penalty) [15], позволяющая исключить доминирование легких совпадений, которые влияют на метрики Rank и mAP. Для вычисления mINP вводятся дополнительные метрики: NP (Negative Penalty) – отрицательный штраф, назначаемый за неверные предсказания для i -го запроса и уменьшающий вероятность правильной реидентификации при неправильном нахождении самого сложного совпадения; INP (Inverse Negative Penalty) – обратная величина для NP, рост которой свидетельствует об улучшении производительности системы. Метрика mINP характеризует среднее значение INP для всех запросов и вычисляется по формуле

$$mINP = \frac{1}{Q} \sum_i (1 - NP_i) = \frac{1}{Q} \sum_i \left(1 - \frac{R_i^{\text{hard}} - |G_i|}{R_i^{\text{hard}}} \right) = \frac{1}{Q} \sum_i \frac{|G_i|}{R_i^{\text{hard}}}, \quad (6)$$

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

Результаты экспериментальных исследований. Для определения наиболее эффективного метода аугментации данных использовался алгоритм повторной идентификации (URL: https://github.com/layumi/Person_reID_baseline_pytorch), реализованный на фреймворке pyTorch. Оптимальные значения скорости обучения и размера пакета определялись эмпирически, наилучшие показатели точности повторной идентификации для модели получены при $lr = 0,07$ и $B = 16$. Применяемый алгоритм повторной идентификации включает метод регуляризации dropout [6] и такие преобразования, как заполнение по границам изображения (padding), отражение по горизонтали (random horizontal flip) и кадрирование (random crop). В процессе обучения используется планировщик скорости, уменьшающий lr в 0,1 раз каждые 40 эпох. Скорость обучения скрытых слоев СНС в 0,1 раза меньше скорости обучения классификационного слоя. Для предложенного подхода в течение первых 45 эпох обучение осуществляется на аугментированных данных, а затем продолжается на исходных изображениях.

Применение аугментированных данных в технологии двухэтапного обучения позволяет повысить точность повторной идентификации по сравнению с базовым подходом и уменьшить временные затраты по сравнению с методом, использующим предварительное обучение на миллионах изображений, обеспечивая при этом сопоставимую точность. Процесс обучения с использованием предложенного подхода может занимать от полутора до шести часов на персональном компьютере с такими основными характеристиками, как Intel Core i5 3.11 GHz, 16 Gb RAM, Nvidia GeForce RTX-3060 6 Gb, в зависимости от набора данных, а предварительное обучение на ImageNet может длиться до нескольких недель (URL: <https://cs231n.github.io/transfer-learning/>). После этого этапа необходимо выполнить точную настройку на наборе данных для повторной идентификации.

В табл. 1 рассматриваются примеры предсказаний алгоритма повторной идентификации с использованием СНС ResNet-50, обученной с помощью различных методов аугментации для двух запросов. Запрос обозначен как query, ошибочные предсказания – false, верные – true.

Анализ данных табл. 1 показал, что для первого запроса при использовании СНС, обученной без аугментации, не было ни одного верного ответа среди пяти первых вариантов предсказаний. При тестировании СНС, для обучения которой применялись методы аугментации random erasing, Cutmix, MixUp и MST, было допущено от двух до четырех ошибок. При использовании предложенного подхода, включающего технологию двухэтапного обучения и аугментацию, была допущена лишь одна ошибка для первого запроса.

Для второго запроса при применении для обучения предложенного подхода не было ни одного ошибочного прогноза. При обучении СНС с аугментацией данных методами random erasing и MST при тестировании получены четыре верных предсказания среди первых пяти результатов. СНС, для обучения которой применялись методы аугментации MixUp и CutMix, допущены две и три ошибки соответственно. Модель, обученная без аугментации, позволила получить одно верное предсказание.















Необходимое количество эпох обучения для обеспечения наибольшей эффективности реидентификации определялось экспериментально. Модель обучалась на протяжении 110 эпох с использованием двухэтапного обучения и предлагаемого метода аугментации. Для каждой пятой эпохи определялась точность работы алгоритма повторной идентификации в метриках Rank1, mAP и mINP. На рис. 5 показаны графики зависимости точности реидентификации от количества эпох обучения с использованием СНС ResNet-50 и набора данных Market-1501.

На графиках, представленных на рис. 4, видно, что наибольшая точность для двухэтапного обучения с применением предлагаемого метода аугментации достигается на 80-й эпохе: Rank1 = 0,8815, mAP = 0,7025, mINP = 0,3506.

Результаты работы алгоритма повторной идентификации людей с использованием ResNet-50, для обучения которой применялись различные методы аугментации

Table 1

The results of the work of the algorithm for person re-identification using ResNet-50, trained by various augmentation methods

Метод аугментации <i>Augmentation method</i>	Запрос 1 <i>Query 1</i>	Результат реидентификации для запроса 1 <i>Re-identification result for query 1</i>	Запрос 2 <i>Query 2</i>	Результат реидентификации для запроса 2 <i>Re-identification result for query 2</i>
Без аугментации <i>Without augmentation</i>				
Random erasing				
Cutmix				
MixUP				
MST				
Предлагаемый подход <i>Proposed approach</i>				

Для определения эффективности применения двухэтапного обучения по сравнению с одноэтапным СНС ResNet-50 обучалась на наборе данных Market-1501 с использованием предложенного метода аугментации на протяжении 80 эпох. Была получена точность Rank1 = 0,8708, mAP = 0,6886, mINP = 0,3401. Это подтверждает, что применение аугментации только на предварительном этапе более эффективно, чем ее применение на протяжении всего процесса обучения.

Для сравнения предлагаемой технологии двухэтапного обучения и метода аугментации с существующими методами, использующими расширение обучающей выборки, применялись СНС ResNet-50 [16] и DenseNet-121 [17]. Для обучения и тестирования использовались наборы данных Market-1501 [18], Duke-MTMC [19] и MSMT17 [20]. Результаты экспериментов представлены в табл. 2.

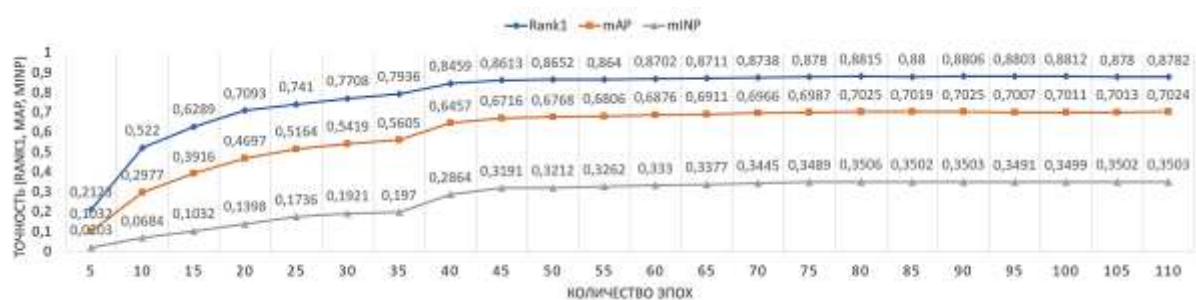


Рис. 5. Зависимость точности повторной идентификации людей в метриках Rank1, mAP и mINP от количества эпох обучения при использовании CNN ResNet-50 и набора данных Market-1501

Fig. 5. Dependence of person re-identification accuracy in the metrics Rank1, mAP and mINP on the training epochs number using CNN ResNet-50 and the Market-1501 dataset

Таблица 2

Сравнение точности методов аугментации для CNN ResNet-50 и DenseNet-121 на наборах данных Market-1501, Duke-MTMC и MSMT17

Table 2

Comparison of the accuracy of augmentation methods for CNN ResNet-50 and DenseNet-121 on Market-1501, Duke-MTMC and MSMT17 datasets

Метод Method	Метрики Metrics	Набор данных для обучения и тестирования Dataset for train and test					
		Market-1501		Duke-MTMC		MSMT17	
		ResNet-50	DenseNet-121	ResNet-50	DenseNet-121	ResNet-50	DenseNet-121
Без аугментации Without augmentation	Rank1	0,8319	0,8376	0,7285	0,7280	0,4954	0,5498
	mAP	0,6108	0,6248	0,5253	0,5316	0,2483	0,2667
	mINP	0,2395	0,2686	0,1453	0,1588	0,251	0,261
Random erasing [8]	Rank1	0,8453	0,8424	0,7361	0,7437	0,4931	0,5476
	mAP	0,6508	0,6490	0,5506	0,5634	0,2469	0,2881
	mINP	0,2773	0,2822	0,1650	0,1775	0,240	0,320
CutMix [9]	Rank1	0,8622	0,8548	0,7424	0,7459	0,4951	0,5500
	mAP	0,6645	0,6632	0,5610	0,5702	0,2562	0,2885
	mINP	0,2792	0,2897	0,1738	0,1778	0,279	0,3185
MixUP [11]	Rank1	0,8480	0,8521	0,7365	0,7428	0,5265	0,5820
	mAP	0,6402	0,6482	0,5401	0,5509	0,2585	0,2985
	mINP	0,2601	0,2802	0,1539	0,1688	0,237	0,310
MST [10]	Rank1	0,8607	0,8607	0,7478	0,7531	0,5071	0,5625
	mAP	0,6757	0,6831	0,5872	0,5911	0,2683	0,3035
	mINP	0,3064	0,3218	0,1957	0,1931	0,318	0,372
Предлагаемый подход Proposed approach	Rank1	0,8815	0,8732	0,7590	0,7666	0,6022	0,6116
	mAP	0,7025	0,6916	0,5823	0,5867	0,3259	0,3366
	mINP	0,3506	0,3431	0,2022	0,1980	0,3950	0,4170

Из данных табл. 2 следует, что предлагаемый подход позволяет получить наилучшие результаты (выделены жирным шрифтом) точности повторной идентификации практически во всех проведенных экспериментах при сравнении с другими методами аугментации. Только для набора данных Duke-MTMC он незначительно уступает методу аугментации MST в метрике mAP для CNN ResNet-50 и DenseNet-121. Наибольшая эффективность предлагаемого подхода отмечается на наборе данных MSMT, так как для ResNet-50 увеличена точность по сравнению с моделью, обученной без аугментации, в метрике Rank1 с 0,4954 до 0,6022, в mAP с 0,2483 до 0,3259 и mINP с 0,251 до 0,395.

Для набора данных Market-1501 значительное увеличение достигнуто по всем трем метрикам оценки точности для СНС ResNet-50: для Rank1 с 0,8319 до 0,8815; mAP с 0,6108 до 0,7025; mINP с 0,2395 до 0,3506.

При сравнении приведенных в табл. 2 результатов экспериментов со значениями, представленными на графике (см. рис. 5), видно, что уже на 50-й эпохе при использовании двухэтапного обучения и предложенного метода аугментации модель достигает значений Rank1 = 0,8652 и mAP = 0,6768. Это превосходит точность СНС, обученной в течение 80 эпох с CutMix (Rank1 = 0,8622, mAP = 0,6645) и MST (Rank1 = 0,8607, mAP = 0,6757). В метрике mINP предлагаемый метод аугментации уже на 45-й эпохе обучения продемонстрировал лучший показатель по сравнению с другими, при которых СНС обучалась на протяжении 80 эпох. Предлагаемый подход подразумевает, что аугментация применяется только на первых 45 эпохах. Таким образом, использование предложенного метода аугментации оказывается более эффективным как для СНС, обученной без расширения обучающей выборки, так и для модели, обученной с использованием других методов аугментации.

Заключение. В работе предложен подход для повышения точности повторной идентификации людей, включающий двухэтапное обучение СНС и аугментацию данных. Технология двухэтапного обучения основана на концепции переноса обучения (transfer learning) и отличается от классического подхода, предполагающего использование для предобучения больших наборов данных, применением на данном этапе наборов, расширенных за счет аугментации, а для точной настройки весовых коэффициентов – применением для обучения изображений из исходного набора. Преимущество предложенной технологии заключается в том, что из-за невысокой вычислительной сложности для аугментации не потребуются значительные вычислительные затраты на выполнение предварительного обучения СНС. Использование на разных этапах обучения различных данных не позволяет СНС запоминать обучающие примеры, тем самым предотвращая переобучение.

В основе предложенного метода аугментации лежат циклический сдвиг пикселей изображения, исключение цветности и замещение фрагмента уменьшенной копией другого изображения. Такой метод аугментации позволяет получить большое разнообразие обучающих данных, что увеличивает робастность СНС к окклюзиям, освещенности, низкому разрешению изображения, зависимости от местоположения отличительных признаков.

Применение технологии двухэтапного обучения и предложенного метода аугментации данных для СНС ResNet-50, обученной на разных наборах данных, позволило повысить точность Rank1 на 4,18–21,55 %, mAP на 10,85–31,25 %, mINP на 39,16–57,37 % по сравнению с СНС, обученной без аугментации. Для СНС DenseNet-121 увеличение точности алгоритма реидентификации людей составило для Rank1 4,25–11,24 %, mAP 10,36–26,21 %, mINP 24,69–60 % при разных наборах данных.

Вклад авторов. С. А. Игнатьева разработала технологию двухэтапного обучения и метод аугментации данных, провела экспериментальные исследования. Р. П. Богуш сформулировал цель и задачи исследования, выполнил анализ и интерпретацию полученных результатов.

Список использованных источников

1. ImageNet: A large-scale hierarchical image database / J. Deng [et al.] // 2009 IEEE Conf. on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009. – Miami, 2009. – P. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
2. Unsupervised pre-training for person re-identification / D. Fu [et al.] // 2021 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021. – Nashville, 2021. – P. 14745–14754. <https://doi.org/10.1109/CVPR46437.2021.01451>
3. Богатырева, А. А. Исследование способности к transfer learning сверточных нейронных сетей, обученных на ImageNet / А. А. Богатырева, А. Р. Виноградова, С. А. Тихомирова // Междунар. журнал прикладных и фундаментальных исследований. – 2019. – № 7. – С. 106–111.

4. Конарев, Д. И. Повышение точности предварительно обученных нейронных сетей путем тонкой настройки / Д. И. Конарев, А. А. Гуламов // Материалы конф. «Информационные технологии в управлении», Санкт-Петербург, 6–8 окт. 2020 г. – СПб., 2020. – С. 200–212.
5. DeVries, T. Improved Regularization of Convolutional Neural Networks with CutOut / T. DeVries, G. W. Taylor. – 2017. – Mode of access: <https://doi.org/10.48550/arXiv.1708.04552>. – Date of access: 09.08.2022.
6. Dropout: A simple way to prevent neural networks from overfitting / N. Srivastava [et al.] // *J. of Machine Learning Research*. – 2014. – No. 15. – P. 1929–1958. <https://doi.org/10.5555/2627435.2670313>
7. Choice of activation function in convolution neural network for person re-identification in video surveillance systems / H. Chen [et al.] // *Programming and Computer Software*. – 2022. – Vol. 48, no. 5. – P. 312–321. <http://doi.org/10.1134/S0361768822050036>
8. Random Erasing Data Augmentation / Z. Zhong [et al.]. – 2020. – Mode of access: <https://doi.org/10.1609/AAAI.V34I07.7000>. – Date of access: 09.08.2022.
9. CutMix: Regularization strategy to train strong classifiers with localizable features / S. Yun [et al.] // 2019 IEEE/CVF Intern. Conf. on Computer Vision (ICCV), Seoul, Korea (South), 27 Oct. – 2 Nov. 2019. – Seoul, 2019. – P. 6022–6031. <https://doi.org/10.1109/ICCV.2019.00612>
10. Cut-thumbnail: A novel data augmentation for convolutional neural network / T. Xie [et al.] // *Proc. of the 29th ACM Intern. Conf. on Multimedia, Virtual Event, China, 20–24 Oct. 2021. – Virtual Event, China, 2021. – P. 1627–1635. <https://doi.org/10.1145/3474085.3475302>*
11. Mixup: Beyond Empirical Risk Minimization / H. Zhang [et al.]. – 2018. – Mode of access: <https://doi.org/10.48550/arXiv.1710.09412>. – Date of access: 09.08.2022.
12. ImageNet-Trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness / R. Geirhos [et al.]. – 2019. – Mode of access: <https://doi.org/10.48550/arXiv.1811.12231>. – Date of access: 09.08.2022.
13. Gong, Y. An Effective Data Augmentation for Person Re-identification / Y. Gong, Z. Zeng. – 2021. – Mode of access: <https://doi.org/10.48550/arXiv.2101.08533>. – Date of access: 09.08.2022.
14. Adversarially occluded samples for person re-identification / H. Huang [et al.] // 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. – Salt Lake City, 2018. – P. 5098–5107. <https://doi.org/10.1109/CVPR.2018.00535>
15. Deep learning for person re-identification: A survey and outlook / M. Ye [et al.] // *IEEE Transactions on Pattern Analysis and Machine Intelligence*. – 2021. – Vol. 44, iss. 6. – P. 2872–2893. <https://doi.org/10.1109/TPAMI.2021.3054775>
16. Deep residual learning for image recognition / K. He [et al.] // 2016 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016. – Las Vegas, 2016. – P. 770–778. <https://doi.org/10.1109/cvpr.2016.90>
17. Huang, G. Densely connected convolutional networks / G. Huang, Z. Liu, K. Q. Weinberger // 2017 IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. – Honolulu, 2017. – P. 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
18. Scalable person re-identification: A benchmark / L. Zheng [et al.] // 2015 IEEE Intern. Conf. on Computer Vision (ICCV), Santiago, Chile, 7–13 Dec. 2015. – Santiago, 2015. – P. 1116–1124. <https://doi.org/10.1109/ICCV.2015.133>
19. Performance Measures and a Data Set for Multi-target, Multi-camera Tracking / E. Ristani [et al.]. – 2016. – Mode of access: https://doi.org/10.1007/978-3-319-48881-3_2. – Date of access: 09.08.2022.
20. Person transfer GAN to bridge domain gap for person re-identification / L. Wei [et al.] // 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018. – Salt Lake City, 2018. – P. 79–88. <https://doi.org/10.1109/CVPR.2018.00016>

References

1. Deng J., Dong W., Socher R., Li L., ..., Fei-Fei L. ImageNet: A large-scale hierarchical image database. *2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009*. Miami, 2009, pp. 248–255. <https://doi.org/10.1109/CVPR.2009.5206848>
2. Fu D., Chen D., Bao J., Yang H., Yuan L., ..., Chen D. Unsupervised pre-training for person re-identification. *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021*. Nashville, 2021, pp. 14745–14754. <https://doi.org/10.1109/CVPR46437.2021.01451>

3. Bogatyreva A. A., Vinogradova A. R., Tikhomirova S. A. *Study of the ability of convolution neural networks pretrained on ImageNet to transfer learning*. *Mezhdunarodnyj zhurnal prikladnyh i fundamental'nyh issledovanij [International Journal of Applied and Fundamental Research]*, 2019, no. 7, pp. 106–111 (In Russ.).
4. Konarev D. I., Gulamov A. A. *Improving the accuracy of pretrained neural networks by fine-tuning*. *Materialy konferencii "Informacionnye tehnologii v upravlenii"*, Sankt-Peterburg, 6–8 oktjabrja 2020 g. [*Proceedings of the Conference "Information Technologies in Management", Saint Petersburg, 6–8 October 2020*]. Saint Petersburg, 2020, pp. 200–212 (In Russ.).
5. DeVries T., Taylor G. W. *Improved Regularization of Convolutional Neural Networks with CutOut*, 2017. Available at: <https://doi.org/10.48550/arXiv.1708.04552> (accessed 09.08.2022).
6. Srivastava N., Hinton G. E., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014, no. 15, pp. 1929–1958. <https://doi.org/10.5555/2627435.2670313>
7. Chen H., Ihnatsyeva S., Bohush R., Ablameyko S. Choice of activation function in convolution neural network for person re-identification in video surveillance systems. *Programming and Computer Software*, 2022, vol. 48, no. 5, pp. 312–321. <http://doi.org/10.1134/S0361768822050036>
8. Zhong Z., Zheng L., Kang G., Li S., Yang Y. *Random Erasing Data Augmentation*, 2020. Available at: <https://doi.org/10.1609/AAAI.V34I07.7000> (accessed 09.08.2022).
9. Yun S., Han D., Oh S., Chun S., Choe J., Yoo Y. J. CutMix: Regularization strategy to train strong classifiers with localizable features. *2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 27 October – 2 November 2019*. Seoul, 2019, pp. 6022–6031. <https://doi.org/10.1109/ICCV.2019.00612>
10. Xie T., Cheng X., Wang X., Liu M., Deng J., ..., Liu M. Cut-thumbnail: A novel data augmentation for convolutional neural network. *Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, China, 20–24 October 2021*. Virtual Event, China, 2021. <https://doi.org/10.1145/3474085.3475302>
11. Zhang H., Cissé M., Dauphin Y., Lopez-Paz D. *Mixup: Beyond Empirical Risk Minimization*, 2018. Available at: <https://doi.org/10.48550/arXiv.1710.09412> (accessed 09.08.2022).
12. Geirhos R., Rubisch P., Michaelis C., Bethge M., Wichmann F., Brendel W. *ImageNet-Trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness*, 2019. Available at: <https://doi.org/10.48550/arXiv.1811.12231> (accessed 09.08.2022).
13. Gong Y., Zeng Z. *An Effective Data Augmentation for Person Re-identification*, 2021. Available at: <https://doi.org/10.48550/arXiv.2101.08533> (accessed 09.08.2022).
14. Huang H., Li D., Zhang Z., Chen X., Huang K. Adversarially occluded samples for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*. Salt Lake City, 2018, pp. 5098–5107. <https://doi.org/10.1109/CVPR.2018.00535>
15. Ye M., Shen J., Lin G., Xiang T., Shao L., Hoi S. C. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, vol. 44, iss. 6, pp. 2872–2893. <https://doi.org/10.1109/TPAMI.2021.3054775>
16. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016*. Las Vegas, 2016, pp. 770–778. <https://doi.org/10.1109/cvpr.2016.90>
17. Huang G., Liu Z., Weinberger K. Q. Densely connected convolutional networks. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017*. Honolulu, 2017, pp. 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>
18. Zheng L., Shen L., Tian L., Wang S., Wang J., Tian Q. Scalable person re-identification: A benchmark. *2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015*. Santiago, 2015, pp. 1116–1124. <https://doi.org/10.1109/ICCV.2015.133>
19. Ristani E., Solera F., Zou R. S., Cucchiara R., Tomasi C. *Performance Measures and a Data Set for Multi-target, Multi-camera Tracking*, 2016. Available at: https://doi.org/10.1007/978-3-319-48881-3_2 (accessed 09.08.2022).
20. Wei L., Zhang S., Gao W., Tian Q. Person transfer GAN to bridge domain gap for person re-identification. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018*. Salt Lake City, 2018, pp. 79–88. <https://doi.org/10.1109/CVPR.2018.00016>

Информация об авторах

Игнатъева Светлана Александровна, магистр технических наук, аспирант кафедры вычислительных систем и сетей, Полоцкий государственный университет имени Евфросинии Полоцкой.

E-mail: s.ignatieva@pdu.by

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

Богущ Рихард Петрович, доктор технических наук, доцент, заведующий кафедрой вычислительных систем и сетей, Полоцкий государственный университет имени Евфросинии Полоцкой.

E-mail: r.bogush@pdu.by

<https://orcid.org/0000-0002-6609-5810>

Information about the authors

Sviatlana A. Ihnatsyeva, M. Sc. (Eng.), Postgraduate Student of the Department of Computing Systems and Networks, Euphrosyne Polotskaya State University of Polotsk.

E-mail: s.ignatieva@pdu.by

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

Rykhard P. Bohush, D. Sc. (Eng.), Assoc. Prof., Head of the Department of Computing Systems and Networks, Euphrosyne Polotskaya State University of Polotsk.

E-mail: r.bogush@pdu.by

<https://orcid.org/0000-0002-6609-5810>