

ИГНАТЬЕВА С.А., БОГУШ Р.П.

## АЛГОРИТМ РЕИДЕНТИФИКАЦИИ ЛЮДЕЙ ПО ИЗОБРАЖЕНИЯМ СИСТЕМ ВИДЕОНАБЛЮДЕНИЯ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВОГО СОСТАВНОГО ДЕСКРИПТОРА

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

*Для повышения точности реидентификации людей в распределенных системах видеонаблюдения важным является использование алгоритма, обеспечивающего эффективность при перекрытии человека другими людьми или объектами. Поэтому для такой задачи разработан алгоритм, предполагающий формирование составного дескриптора, который включает глобальный вектор признаков изображения человека и три локальных, для его верхней, средней и нижней частей. Выделение областей интереса осуществляется на основе результатов обнаружения ключевых точек изображения тела человека. Если часть изображения человека перекрывается другими людьми или объектами, то она относится к невидимой. Изображение скрытой части человека не используется в формировании локального признака. Для его получения вычисляется усредненное значение таких признаков  $k$ -ближайших соседей изображения человека. Выполненные эксперименты свидетельствуют о повышении точности повторной идентификации для наборов данных Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077.*

**Ключевые слова:** машинное обучение, сверточные нейронные сети, вектор признаков, PolReID1077, аугментация данных

### Введение

Широкое распространение распределённых систем видеонаблюдения позволяет решать множество практических задач, в том числе и для общественной безопасности. Актуальной среди них является реидентификация (повторная идентификация) людей, которая позволяет определить присутствие заданного человека по его изображению на видеоданных в различных местах или в разное время. Для повышения ее эффективности необходимо выявить признаки и построить алгоритм, которые позволят с высокой точностью определить наиболее схожие изображения людей из имеющейся выборки (галереи) с образцом (запросом). Развитие вычислительной техники и методов машинного обучения, включая сверточные нейронные сети (СНС), способствуют возможности значительного увеличения точности решения такой задачи. Однако, при реидентификации людей на изображении по внешнему виду СНС извлекает отличительные характеристики области интереса в целом, а вектор признаков зависит от особенностей фона, видеокамеры, уровня освещенности, наличия схожих объектов на снимаемой сцене.

Одной из наиболее существенных проблем является наличие частичных перекрытий изображения человека другими объектами, т. е. окклюзий. Соответственно, в таких случаях дескриптор описывает объект интереса с учетом признаков скрываемых его объектов, что значительно ухудшает точность реидентификации. Для решения указанной пробле-

мы предлагается применять различные подходы, включая использование аугментации при обучении, направленной на повышение устойчивости СНС к окклюзиям [1, 2], и анализ локальных признаков изображений людей [3-5]. Методы аугментации позволяют улучшить обобщающую способность СНС, но при этом учитываются характеристики всего изображения человека, а его локальные особенности, важные для реидентификации, не всегда могут быть выявлены и использованы.

Горизонтальное разделение изображения человека на части с последующим независимым сравнением с соответствующими им фрагментами изображений других людей предложено в [6]. Основным недостатком такого подхода является то, что человек должен принимать строго вертикальное положение и части тела должны располагаться корректно. Очевидно, что при наличии ошибок генерации ограничительных рамок или окклюзий, точность работы такого алгоритма будет значительно уменьшаться. В [4] объект интереса разделяется на шесть горизонтальных областей и на каждой из них определяются ключевые точки с использованием системы определения позы человека AlphaPose [7]. Если во фрагменте изображения не обнаружены ключевые точки, то он считается невидимым и не учитывается при сравнении на последнем этапе алгоритма. В [8] видимые части тела человека определяются с применением оценщика поз (pose estimator), который позволяет выявить расположение локальных областей и выполнить покомпонентное сопоставление векторов признаков только видимых участков

для запроса и изображений из галереи. Предлагается использовать модель DeepLabV3 для обнаружения головы, верхней и нижней части тела человека, обуви в работе [5]. Для детектированных областей формируются локальные вектора признаков, а скрытые части игнорируются при реидентификации. В [9] повышение точности повторной идентификации основывается на использовании семантического анализа для выделения пяти локальных областей. Сопоставление признаков осуществляется для каждой локальной области с учетом вероятности достоверности сегментации области. Общим недостатком алгоритмов, представленных в [3-5], является то, что не обеспечивается участие скрытых регионов объекта интереса в формировании дескриптора для его описания, что не позволяет в полной мере обеспечить высокую точность повторной идентификации людей в распределенных системах видеонаблюдения. Поэтому целью работы является разработка алгоритма, позволяющего учитывать признаки скрытого фрагмента изображения человека при его реидентификации.

### Формализация алгоритма реидентификации людей

Разработан алгоритм реидентификации людей, который предполагает описание изображения каждого человека составным дескриптором, включающим глобальный вектор признаков всего изображения и три локальных, описывающих верхнюю, среднюю и нижнюю части изображения человека. Принятие решения об их видимости осуществляется на основе результатов обнаружения ключевых точек изображения тела человека. Если человек на изображении перекрывается другими людьми или объектами, то соответствующая область относится к невидимой. Скрытые фрагменты изображения не используются в формировании локальных признаков. Для их получения вычисляется усредненное значение признаков соответствующих частей  $k$ -ближайших соседей изображения человека. Ранжирование вычисленных значений для определения наиболее схожих изображений из галереи с запросом выполняется по метрике косинусного расстояния.

#### Формирование составного дескриптора

На кадре видеопоследовательности с помощью детектора осуществляется обнаружение изображения человека и выделение его ограничительной прямоугольной рамкой. Данная область кадра нормализуется до размеров  $256 \times 128$ , по аналогии с современными существующими подходами [10] и поступает на вход алгоритма реидентификации. Однако она может включать изображения и других людей. Поэтому на первом этапе выполняется проверка количества людей области интереса с учетом возможных перекрытий между ними. Если обнаружено

несколько человек, то объектом интереса, который в дальнейшем анализируется, считается тот, площадь ограничительного прямоугольника которого больше. Затем для изображения человека  $I^{(Y)}$ , где  $Y$  – идентификатор человека, формируется составной нейросетевой дескриптор, который включает глобальный вектор  $f_{global}(I^{(Y)})$  всей входной области и локальные вектора  $f_{p1}(I_{p1}^{(Y)})$ ,  $f_{p2}(I_{p2}^{(Y)})$ ,  $f_{p3}(I_{p3}^{(Y)})$  для верхней  $p1$ , средней  $p2$  и нижней  $p3$  ее частей соответственно, полученные с использованием СНС. Для выделения фрагментов на изображении человека выполняется детектирование ключевых точек, на основе координат которых определяются координаты области. При этом формируются таких же размеров маски  $M_{p1}$ ,  $M_{p2}$ , и  $M_{p3}$ , состоящие из нулей и единиц. Единичной областью в данных масках описываются видимые локальные фрагменты, соответственно, нулевые значения характеризуют скрытые части. Для каждого изображения формируется составной вектор признаков  $f_{gen}^{(Y)} = \{f_{global}(I^{(Y)}), f_{p1}(I_{p1}^{(Y)}), f_{p2}(I_{p2}^{(Y)}), f_{p3}(I_{p3}^{(Y)})\}$ , где  $I_{p1}^{(Y)} = (I^{(Y)} \circ M_{p1})$ ,  $I_{p2}^{(Y)} = (I^{(Y)} \circ M_{p2})$ ,  $I_{p3}^{(Y)} = (I^{(Y)} \circ M_{p3})$ ;  $\circ$  – поэлементное умножение.

Локализация частей изображения человека  $p1$ ,  $p2$ ,  $p3$  основана на обнаружении и анализе семнадцати ключевых точек ( $kp_i$ ,  $i$  – порядковый номер точки,  $i \in 1...17$ ) на различных частях изображения тела человека (рисунок 1), включая: нос ( $kp_1$ ), глаза ( $kp_2$ ,  $kp_3$ ), уши ( $kp_4$ ,  $kp_5$ ), плечи ( $kp_6$ ,  $kp_7$ ), локти ( $kp_8$ ,  $kp_9$ ), запястья ( $kp_{10}$ ,  $kp_{11}$ ), бедра ( $kp_{12}$ ,  $kp_{13}$ ), колени ( $kp_{14}$ ,  $kp_{15}$ ) и лодыжки ( $kp_{16}$ ,  $kp_{17}$ ). Точка считается обнаруженной, если  $Tkp_i \geq 0,5$ , где  $Tkp_i$  – степень уверенности нейронной сети в достоверности предсказания. На основе координат ключевых точек и результатов их обнаружения фрагмент  $p_2$  включается как в  $p_1$ , так и в  $p_3$ , т. е. присутствует перекрытие среди фрагментов.



Рисунок 1. Разделение изображение человека на локальные области

### Локализация средней области изображения человека

Средняя часть изображения человека перекрывается верхней и нижней, поэтому ее локализация выполняется в первую очередь. Область будет считаться видимой, а соответствующий компонент вектора признаков принимается действительным, если видна хотя бы одна ключевая точка  $kp_{12}$  или  $kp_{13}$ . Для определения ее местоположения на первом шаге необходимо вычислить ее центр:

$$x_{c_{p2}} = (x_{kp12} - x_{kp13}) / 2; y_{c_{p2}} = (y_{kp12} - y_{kp13}) / 2. \quad (1)$$

Значения координат верхнего левого ( $x_{1_{p2}}; y_{1_{p2}}$ ) и правого нижнего ( $x_{2_{p2}}; y_{2_{p2}}$ ) углов можно определить, как:

$$x_{1_{p2}} = x_{c_{p2}} - w_{p2}/2; y_{1_{p2}} = y_{c_{p2}} - h_{p2}/2;$$

$$x_{2_{p2}} = x_{c_{p2}} + w_{p2}/2; y_{2_{p2}} = y_{c_{p2}} + h_{p2}/2;$$

где  $w_{p2}$  и  $h_{p2}$  – ширина и высота области  $p_2$ .

В области бедер могут быть расположены руки, поэтому для уменьшения их влияния ширина данной области выбирается меньше ширины изображения. При  $H = 256$  и шириной  $W = 128$  экспериментально определено усредненное значение  $w_{p2} = 85$ , а высота  $h_{p2} = 45$  пикселей.

Если локализована только одна ключевая точка  $kp_{12}$  или  $kp_{13}$ , то ее координаты принимаются как  $x_{c_{p2}}$  и  $y_{c_{p2}}$ .

### Локализация верхней области изображения человека

Фрагмент  $p_1$  будет считаться видимым, если обнаружено минимум по одной ключевой точке на любых четырех частях тела из следующих: лицо, плечо, локоть, запястье, бедро. В общем случае, определение верхнего левого ( $x_{1_{p1}}; y_{1_{p1}}$ ) и нижнего правого ( $x_{2_{p1}}; y_{2_{p1}}$ ) углов верхней части  $p_1$  изображения человека выполняется на основе  $p_2$  и расположения головы  $p_{head}$  при условии их видимости. Область  $p_{head}$  считается невидимой, если ни одной ключевой точки из ( $kp_1 - kp_5$ ) не обнаружено.

Если  $p_{head}$  не перекрыта другими объектами, то чтобы получить координаты ограничительного прямоугольника, описывающего голову человека, анализируются первые пять ключевых точек ( $kp_1 - kp_5$ ). Среди них определяются две с наибольшим расстоянием между ними  $d_{kp_{head}}$ , которое необходимо для вычисления высоты  $h_{p_{head}}$  и ширины  $w_{p_{head}}$  области  $p_{head}$ :

$$h_{p_{head}} = (0,9 \cdot d_{kp_{head}}) \cdot 2; w_{p_{head}} = (0,9 \cdot d_{kp_{head}}) \cdot 2,$$

а также координат ( $x_{c_{head}}; y_{c_{head}}$ ) центра области  $p_{head}$  по аналогии с (1).

Если обнаружена только одна точка из ( $kp_1 - kp_5$ ), то для принятого изображения размером  $256 \times 128$   $h_{p_{head}} = w_{p_{head}} = 40$  пикселей, и координаты центра области принимаются равными координатам единственной обнаруженной ключевой точки. Коор-

динаты верхнего левого ( $x_{1_{head}}; y_{1_{head}}$ ) и правого нижнего ( $x_{2_{head}}; y_{2_{head}}$ ) углов определяются как:

$$x_{1_{head}} = x_{c_{head}} - w_{head}/2; y_{1_{head}} = y_{c_{head}} - h_{head}/2;$$

$$x_{2_{head}} = x_{c_{head}} + w_{head}/2; y_{2_{head}} = y_{c_{head}} + h_{head}/2;$$

Высота  $h_{p1}$  области  $p1$  определяется верхней границей  $p_{head}$  и нижней области  $p2$ , а ширина соответствует  $W$ . Таким образом  $x_{1_{p1}} = 0$ ,  $y_{1_{p1}} = y_{1_{head}}$ ,  $x_{2_{p1}} = W$  и  $y_{2_{p1}} = y_{2_{p2}}$ .

Если область  $p_{head}$  считается невидимой, то  $h_{p1} = H/2$ , и  $y_{1_{p1}} = y_{1_{head}} - h_{p1}$ , а в случае скрытого фрагмента  $p_2$  используется правило  $y_{2_{p1}} = y_{1_{head}} - h_{p1}$ .

### Локализация нижней области изображения человека

Нижняя область фигуры человека  $p_3$  является видимой при обнаружении хотя бы по одной ключевой точки на двух частях тела, таких как бедра, колени или лодыжки. Определение координат области  $p_3$  основано на использовании данных локальных фрагментов стопы  $p_{foot1}$  или  $p_{foot2}$ , а также области  $p2$ . Положение  $p_{foot1}$  и  $p_{foot2}$  на изображении зависит от предсказанных значений  $kp_{16}$  и  $kp_{17}$ . Координаты верхнего левого ( $x_{1_{foot1}}; y_{1_{foot1}}$ ) и правого нижнего ( $x_{2_{foot1}}; y_{2_{foot1}}$ ) углов для области обеих стоп вычисляются как:

$$x_{1_{foot1}} = x_{kp_{foot1}} - w_{foot1}/2; y_{1_{foot1}} = y_{kp_{foot1}} - h_{foot1}/2;$$

$$x_{2_{foot1}} = x_{kp_{foot1}} + w_{foot1}/2; y_{2_{foot1}} = y_{kp_{foot1}} + h_{foot1}/2;$$

В среднем стопа занимает около 20 пикселей для изображения заданного размера, но в связи с тем, что для предсказания положения стопы используется всего одна ключевая точка, которая не позволяет определить величину и направление поворота стопы, то по четырем основным направлениям учитывается по 20 пикселей. При  $H = 256$  и  $W = 128$  примем  $w_{foot} = h_{foot} = 40$ .

Координаты верхнего левого и нижнего правого угла единичной области маски  $M_{p3}$  могут быть определены как:  $x_{1_{p3}} = 0$ ,  $y_{1_{p3}} = y_{1_{p2}}$ ,  $x_{2_{p3}} = W$  и  $y_{2_{p3}} = \max(y_{2_{foot1}}, y_{2_{foot2}})$ . Если ключевые точки  $kp_{16}$  и  $kp_{17}$  не были обнаружены, то  $y_{2_{p3}} = y_{1_{p3}} + h_p$ , а в случае, если они не детектированы  $kp_{12}$ ,  $kp_{13}$ , то  $y_{2_{p3}} = \max(y_{2_{foot1}}, y_{2_{foot2}}) - h_p$ .

### Замещение признаков для невидимых областей

При первом проходе изображений по сети формируется составной дескриптор для всех изображений галереи. Если одна или несколько локальных областей считаются другими объектами, то соответствующий компонент составного дескриптора будет недействительным и равным нулю. Для изображений, вектор признаков которых имеет нулевые составляющие, выполняется ранжирование таблицы признаков по метрике косинусного подобию для каждого видимого локального фрагмента, для которого определяются  $k$ -ближайших соседей и выбираются  $k_1$ -лучших. Таким образом выявляются какие изображения человека являлись ближайшими для



большого числа действительных компонентов составного вектора признаков. Усредненное значение соответствующей компоненты  $f_p^k(I_p^{(v)})$  вектора признаков  $k_1$ -ближайших соседей используется для замещения недействительного компонента вектора признаков. Значения для  $k$  и  $k_1$  определялись экспериментально, наилучшие показатели точности реидентификации получены при  $k = 8$ ,  $k_1 = 2$ . После замещения недействительных компонент составного дескриптора таблица признаков обновляется для векторов признаков с замещенными фрагментами, и ранжирование выполняется в соответствии с запросом.

### Результаты экспериментов

Проверка количества людей в заданной области, выделение ключевых точек человека и получения координат локальных областей выполнено на основе СНС YOLOv7pose. Для формирования нейросетевых признаков использованы СНС ResNet-50 и DenseNet-121. На рисунке 2 представлены примеры тепловых карт, позволяющие визуализировать обла-

сти изображения, оказывающие большее влияние на принятие сетью решения при использовании признаков с третьего и четвертого слоев ResNet-50. Из рисунка 2а очевидно, что после третьего слоя СНС позволяет выделить три локальные области, каждая из которых характеризует части изображения человека, а после четвертого слоя сети (рисунк 2б), выделяется всего две области, причем одна из них характеризует фон, а не изображение человека. Выполнено тестирование на наборе данных Market-1501 для предобученной ResNet-50: точность реидентификации для вектора признаков после третьего уровня сети в метриках Rank1 = 90.17 и mAP=76.95, а после четвертого слоя получены значения Rank1 = 87.98, mAP=73.71. Поэтому классификационный слой встраивается после третьего уровня сети, уменьшая при этом количество сверточных слоев, но увеличивая точность реидентификации. Следует отметить, что проведение аналогичного эксперимента для DenseNet-121 показало, что эта СНС имеет большую эффективность при применении традиционной архитектуры.

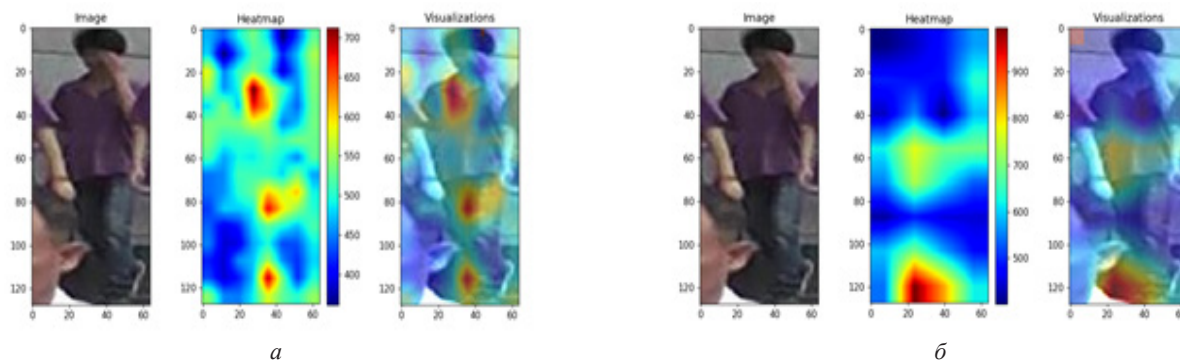


Рисунок 2. Тепловые карты и результат их наложения на изображение полученные на основании вектора признаков: а – с третьего слоя СНС ResNet-50, б – с четвертого слоя СНС ResNet-50

Таблица 1

Результаты оценки точность повторной идентификации людей при использовании разных алгоритмов

Подход к обучению	Метрики	Тестовая выборка, СНС							
		Market-1501		DukeMT-MC-ReID		MSMT17		PolReID1077	
		ResNet-50	DenseNet-121	ResNet-50	DenseNet-121	ResNet-50	DenseNet-121	ResNet-50	DenseNet-121
Базовый алгоритм [10] без предобучения	Rank1 mAP	83,19 61,08	83,76 62,48	72,85 52,53	72,80 53,16	49,54 24,83	54,98 26,67	88,94 65,58	90,33 68,39
Предлагаемый алгоритм	Rank1 mAP	86,67 69,49	85,48 70,73	77,83 59,33	76,48 57,70	54,87 28,31	55,23 27,12	89,29 65,27	90,62 68,89
Предлагаемый алгоритм и аугментация [1]	Rank1 mAP	90,26 76,38	87,08 71,38	80,97 65,2	79,26 61,95	63,05 36,08	60,05 30,64	93,83 76,50	94,27 78,11
Предлагаемый алгоритм, аугментация и предобученная СНС	Rank1 mAP	93,02 83,25	93,29 82,26	84,92 71,34	84,38 69,70	74,85 48,89	61,40 32,10	95,85 83,82	97,14 86,03
Предлагаемый алгоритм, аугментация и предобученная СНС, объединенный набор данных	Rank1 mAP	93,17 83,70	94,12 85,73	84,47 72,22	87,79 76,68	76,18 44,62	80,32 57,55	97,03 88,04	97,19 89,32

Результаты исследований по определению точности реидентификации людей при использовании разных подходов на базах данных Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077 представлены в таблице.

Анализ таблицы 1 свидетельствует, что применение предлагаемого алгоритма позволило повысить точность реидентификации относительно базового [10] для всех исследуемых наборов данных и СНС ResNet-50 и DenseNet-121. Для набора данных Market-1501 увеличивается точность в метрике Rank1 на 12 %, в метрике mAP на 37 %, для Duke-MTMC-ReID Rank1 – на 16 % – 21 %, mAP на 37 % – 44 %, для MSMT17 – в метрике Rank1 46 % – 54 %, в метрике mAP 80 % – 116 %, при использовании PolReID1077 – 8 % – 19 % для Rank1, 31 % – 34 % для mAP.

## Заключение

Предложен эффективный алгоритм реидентификации людей для распределенных систем видеонаблюдения. Использование локальных признаков позволяет уменьшить влияние наличия на изображении человека окклюзий за счет формирования для скрытых областей усредненных признаков на основе алгоритма для  $k$ -ближайших соседей из изображений людей, представленных галереи, наиболее схожих с запросом. Выполненные эксперименты с использованием СНС ResNet50 и DenseNet-121 при формировании нейросетевых признаков подтвердили эффективность разработанного алгоритма. Для всех исследуемых наборов данных Market-1501, DukeMTMC-ReID, MSMT17 и PolReID1077 и СНС получены более высокие значения метрик Rank1 и mAP.

## ЛИТЕРАТУРА

1. **Игнатъева, С.А.** Увеличение точности реидентификации людей на основе двухэтапного обучения сверточных нейронных сетей и аугментации / С.А. Игнатъева, Р.П. Богуш // Информатика. – 2023. – Т. 20. – № 1. – С. 40-54.
2. **Zhong, Z.** Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang // AAAI Conference on Artificial Intelligence. – 2017.
3. **Wang, G.** High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification / G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S. Wang, G. Yu, E. Zhou, J. Sun // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – P. 6448-6457.
4. **Yang, J.** Learning to Know Where to See: A Visibility-Aware Approach for Occluded Person Re-identification / J. Yang, J. Zhang, F. Yu, X. Jiang, M. Zhang, X. Sun, Y. Chen, W. Zheng // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – P. 11865-11874.
5. **Yang, Q.** Focus on the Visible Regions: Semantic-Guided Alignment Model for Occluded Person Re-Identification / Q. Yang, P. Wang, Z. Fang, Q. Lu. // Sensors (Basel, Switzerland), Vol. 20, 2020.
6. **Sun, Y.** Beyond Part Models: Person Retrieval with Refined Part Pooling / Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang // European Conference on Computer Vision, 2017.
7. **Fang, H.** RMPE: Regional Multi-person Pose Estimation / H. Fang, S. Xie, Y. Tai, C. Lu. // Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017. – P. 2353-2362.
8. **Miao, J.** Pose-Guided Feature Alignment for Occluded Person Re-Identification / J. Miao, Y. Wu, P. Liu, Y. Ding, Y. Yang // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – P. 542-551.
9. **Kalayeh, M.M.** Human Semantic Parsing for Person Re-identification / M.M. Kalayeh, E. Basaran, M. Gokmen, M.E. Kamasak, M. Shah // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018. – P. 1062-1071.
10. **GitHub** [Электронный ресурс]. – Режим доступа: [https://github.com/layumi/Person\\_reID\\_baseline\\_pytorch](https://github.com/layumi/Person_reID_baseline_pytorch) – Дата доступа 04.09.2023.

## REFERENCES

1. **Ihnatsyeva, S.A.** Uvelichenie tochnosti reidentifikacii ljudej na osnove dvuhjetapnogo obuchenija svertochnyh nejronnyh setej i augmentacii / S.A. Ihnatsyeva, R.P. Bohush // Informatika. – 2023. – Vol. 20. – № 1. – Pp. 40-54.
2. **Zhong, Z.** Random Erasing Data Augmentation / Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang // AAAI Conference on Artificial Intelligence. – 2017.
3. **Wang, G.** High-Order Information Matters: Learning Relation and Topology for Occluded Person Re-Identification / G. Wang, S. Yang, H. Liu, Z. Wang, Y. Yang, S., Wang, G. Yu, E. Zhou, J. Sun // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). – 2020. – Pp. 6448-6457.
4. **Yang, J.** Learning to Know Where to See: A Visibility-Aware Approach for Occluded Person Re-identification / J. Yang, J. Zhang, F. Yu, X. Jiang, M. Zhang, X., Sun, Y. Chen, W. Zheng // 2021 IEEE/CVF International Conference on Computer Vision (ICCV). – 2021. – Pp. 11865-11874.
5. **Yang, Q.** Focus on the Visible Regions: Semantic-Guided Alignment Model for Occluded Person Re-Identification / Q. Yang, P. Wang, Z. Fang, Q. Lu. // Sensors (Basel, Switzerland), Vol. 20, 2020.

6. Sun, Y. Beyond Part Models: Person Retrieval with Refined Part Pooling / Y. Sun, L. Zheng, Y. Yang, Q. Tian, S. Wang // European Conference on Computer Vision, 2017.
7. Fang, H. RMPE: Regional Multi-person Pose Estimation / H. Fang, S. Xie, Y. Tai, C. Lu // Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017. – Pp. 2353-2362.
8. Miao, J. Pose-Guided Feature Alignment for Occluded Person Re-Identification / J. Miao, Y. Wu, P. Liu, Y. Ding, Y. Yang // 2019 IEEE/CVF International Conference on Computer Vision (ICCV). – 2019. – Pp. 542-551.
9. Kalayeh, M.M. Human Semantic Parsing for Person Re-identification / M.M. Kalayeh, E. Basaran, M. Gokmen, M.E. Kamasak, M. Shah // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. – 2018. – Pp. 1062-1071.
10. **GitHub** [Электронный ресурс]. – Режим доступа: [https://github.com/layumi/Person\\_reID\\_baseline\\_pytorch](https://github.com/layumi/Person_reID_baseline_pytorch) – Дата доступа 04.09.2023.

*IHNATSYEVA S.A., BOHUSH R.P.*

## PERSON RE-IDENTIFICATION ALGORITHM BY IMAGE FROM VIDEO SURVEILLANCE SYSTEM USING A NEURAL NETWORK COMPOUND DESCRIPTOR

*Euphrosyne Polotskaya State University of Polotsk  
Polotsk, Republic of Belarus*

*For people re-identification in distributed video surveillance systems, it is important to have an algorithm that ensures efficiency when a person is blocked by other people or objects. Therefore, for this task, an algorithm has been developed that involves the compound descriptor formation, which includes a global features vector of a person's image and three local ones for its upper, middle and down parts. The regions of interest selection is carried out based on the detecting key points results in the image of the human body. If a person's image part is occluded by other people or objects, then it is classified as invisible. The hidden part person image is not used in the formation of the local feature. To obtain it, the average value of the  $k$ -nearest neighbors such features of a person's image is calculated. The experiments performed indicate an increase in re-identification accuracy for the Market-1501, DukeMTMC-ReID, MSMT17 and PolReID1077 datasets.*

**Keywords:** machine learning, convolutional neural networks, feature vector, PolReID1077, data augmentation



**Игнатъева Светлана Александровна**, старший преподаватель кафедры вычислительных систем и сетей факультета информационных технологий Полоцкого государственного университета им. Евфросинии Полоцкой, магистр технических наук.



**Богущ Рихард Петрович**, доктор технических наук, доцент, заведующий кафедрой вычислительных систем и сетей факультета информационных технологий Полоцкого государственного университета им. Евфросинии Полоцкой.

**E-mail:** r.bogush@psu.by