



Applied Artificial Intelligence

An International Journal

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/uaai20>

Focusing on shared areas for partial person re-identification

Shuren Zhou, Fan Zhang & Wenmin Zou

To cite this article: Shuren Zhou, Fan Zhang & Wenmin Zou (2022) Focusing on shared areas for partial person re-identification, Applied Artificial Intelligence, 36:1, 2031818, DOI: 10.1080/08839514.2022.2031818

To link to this article: <https://doi.org/10.1080/08839514.2022.2031818>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 25 Jan 2022.



Submit your article to this journal [↗](#)



Article views: 829



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 1 View citing articles [↗](#)

Focusing on shared areas for partial person re-identification

Shuren Zhou , Fan Zhang, and Wenmin Zou

School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

ABSTRACT

Person re-identification (Re-ID) can achieve ideal performance based on the prerequisite that the sampling image is complete. However, the whole body cannot be detected because pedestrians may be occluded or are at the edge of the surveillance range in real-world scenarios. Consequently, the image only contains part of the visible information of the pedestrian. When using the standard person re-identification to match the partial image with the complete one, we witness the problem of spatial misalignment and interference caused by missing areas. Hence, we propose a focused shared area model (FSA) for partial re-identification to solve such descriptive problems. We use self-supervised learning to locate the shared area and learn region-level features. In addition, we adopt self-attention mechanism to help the network visualize the important features of the image, thus reducing the influence of the background information. Finally, we verify the effectiveness of our method through experiments on two mainstream datasets: Market-1501, DukeMTMC-reID and two important partial datasets: Partial-REID and Partial-iLIDS.



ARTICLE HISTORY

Received 7 May 2021
Revised 4 January 2022
Accepted 18 January 2022

Introduction

Person re-identification(Re-ID) can be understood as image retrieval at a simple level, specifically, it refers to the retrieval of the same pedestrian under different cameras(Zheng, Yang, and Hauptmann 2016). In recent years, the widespread popularity of surveillance camera equipment and people's high requirements for safety as always also make person re-identification have a very important practical significance, which has attracted more and more researchers (Zhang et al. 2017; Wang et al. 2018b; Yu and Zheng 2020; Wang and Zhang 2020).

However, due to the reality that pedestrians are partially occluded, and the limited range of the camera's shooting range, some images captured only contain the information of the body parts of the pedestrian in Figure 1. Applying the method based on the standard person-re-identification(Li,

CONTACT Shuren Zhou  zsr_hn@163.com  School of Computer and Communication Engineering, Changsha University of Science and Technology, Changsha, China

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



Figure 1. The comparison of partial image and holistic image.

Zhu, and Gong 2018; Zheng et al. 2018; Lin et al. 2019; Liu, Chang, and Shen 2020) to the partial person re-identification is not satisfactory. The previous studies on partial person re-identification mainly focused on completing the effective matching of the whole pedestrian image and partial pedestrian image (Sun et al. 2019; He et al. 2018b). The current researches situation have: 1) Sliding window matching (SWM) which constructs part of the pedestrian image in the query set as a sliding window to slide on the overall pedestrian image to locate the most similar area (Zheng et al. 2015b), but it needs to be traversed. The search efficiency is not high. 2) Deep spatial reconstruction (DSR) can match feature maps without being restricted by different image sizes (He et al. 2018a), but due to the need for one-to-one matching, the GPU's tensor computing power utilization is not enough. 3) The partial matching net (PMN) uses the key points of the human skeleton to align the body regions (Iodice and Mikolajczyk 2018), but requires expensive additional clues and annotations. Moreover, the two main problems faced by partial person-re-identification are spatial dislocation and interference in non-shared areas.

In order to address the above problems, we propose a method of focusing the shared area model (FSA). First, we segment the complete image uniformly and use the self-supervised learning to predefine the label of each region. Then, in order to perceive the shared area, we need to classify and predict the pixel based on the generated label. If the pixel belongs to the region, it will get a higher probability, otherwise, the probability value will be very small. Each area corresponds to a probability map, and all the values in the probability map are added to get the visible score of the area. In the feature extraction part, we use the attention mechanism to learn to generate more important features, and weighted pooling the probability map and the visible score with the learned features to obtain the features of each region, and it can be seen that the area with too small score makes almost no contribution to the subsequent similarity calculation. In the testing phase, we first calculate the

Euclidean distance between each area separately, and then add the distances between all parts to get the overall distance. Similarly, we also apply the two loss functions commonly used in person re-identification tasks: cross-entropy loss and triplet loss.

In summary, the contributions of our work are as follows:

(1) In order to solve the two problems of spatial misalignment and interference in non-shared areas of partial re-identification, we proposed a focused shared area model, which can perceive the shared areas between the complete image and partial image.

(2) Because the background of the pedestrian image is too messy and the scene changes variously, in order to reduce the impact of the background, we adopt a channel attention mechanism to let the network pay more attention to the pedestrian foreground area and collect representative and discriminative features.

(3) The effectiveness of our method is verified on two mainstream holistic datasets and two partial datasets.

Related works

Person re-identification based on local feature

The local feature have also been proved to be helpful for person re-identification, which can improve the accuracy of Re-ID by combining with global features (Zhao et al. 2017b; Kumar et al. 2017; Li et al. 2017; Qian et al. 2018). In the previous research work, manual segmentation is a very common method to extract local features (Wang et al. 2018a; Zheng et al. 2019a). Due to the particularity of human body structure, pictures are often divided into several parts along the vertical direction (head, upper body, lower body, etc.), and finally all local features are integrated into a final representation. For instance, PCB(Sun et al. 2018) framework is to evenly divide pedestrian feature map into six pieces, and then conduct loss training on six feature maps, respectively, to predict the ID. However, this kind of method exists the disadvantage of relatively higher image spatial alignment requirement.

Some methods(Sarfraz et al. 2018; Zhao et al. 2017a; Song et al. 2019; Zhu et al. 2020) adopt the body part detectors or human parsing models instead of bounding boxes to locate arbitrary contours of various body parts accurately. For instance, SPReID(Kalayeh et al. 2018) put forward a parsing model with pixel-level accuracy to generate five probability maps (foreground, head, upper body, lower body, and shoes) with different pre-defined human body regions to calculate more reliable parts representation, and then achieved excellent results on different person re-identification benchmarks.

It is another common practice to align the local characteristics of pedestrians with some prior knowledge, which is mainly the pretrained human pose estimation model and skeleton key point model (Ge et al. 2018; Liu et al. 2018; Miao et al. 2019; Miao, Wu, and Yang 2021; Zhu et al. 2019). Postures are easier to label than human parsing, and there are many different datasets that can be easily generalized. This thesis(Wei et al. 2017) divides the pedestrians image into three parts, namely the head, upper body, and lower body, by using the extracted key points of the human body. Finally, the extracted features merged the global and local features. The thesis produced by Zhao et al. (2017a) generates seven body regions through 14 key points positioning, and then extracts the features of different regions and merges them hierarchically. Unlike Spindle Net, the thesis (Zheng et al. 2019b) first estimated key points of the pedestrian using the pose estimation model, and then used affine transformation to justify the same key points. However, the potential gap between the datasets used for pose estimation and the person re-identification datasets is still a problem.

Partial person re-identification

The holistic person re-identification based on deep learning has made significant research progress and a lot of research results. However, the algorithm based on the holistic person re-identification is no longer applicable for the partial person re-identification. In the partial Re-ID study, the query image is partial, while the gallery image is complete. If the partial image is directly matched with the holistic image, it will lead to spatial dislocation and the interference of non-shared areas. Aiming at the problem of partial person re-identification, a data augmentation method of compound batch erasure is proposed to simulate the occlusions(Yan et al. 2021).

The problem of partial person re-identification was proposed by Zheng et al. (2015b) and the search image was constructed as a sliding window to traverse similar areas. In order to solve the spatial misalignment, He et al. (2018a) proposed DSR to reconstruct the depth space features. Later, they proposed an improved method SFR (He et al. 2018b) based on DSR, which used multi-scale pyramid pooling to enhance the applicability of features to multi-scale images and achieved better results.

All the above methods solve the challenges of partial person re-identification by directly matching the image. Semantic segmentation and pose estimation are also commonly used in partial person re-identification, such as the PMN(Iodice and Mikolajczyk 2018) makes use of the pre-trained human post estimation model to extract the skeleton key point, and calculates the similarity of the shared area based on the shared key points. This type of method can match the shared region of two images more accurately, but

requires additional clues and conditions to assist the matching process. To some extent, the performance of such methods depends on the stability and reliability of the priori model and the cost is relatively high.

Attention mechanism

Attention mechanism has been commonly used in the field of natural language processing and computer vision(Ji et al. 2020; Li, Zhu, and Gong 2018; Si et al. 2018; Song et al. 2018), and has achieved great success in the former. In the study of computer vision, the attention mechanism is used to further process visual information. It can help the model assign different weights to each part of the input, extract more critical and important information, and enable the model to make more accurate judgments without incurring greater costs for the calculation and storage of the model. This is also the reason why attention mechanism is so popular. The SENet(Hu, Shen, and Sun 2018) introduced a channel attention mechanism to pay attention to the importance of each channel. Inspired by channel attention and spatial attention, a kind of fused attention CBAM(Woo et al. 2018) is proposed to improve the attention mechanism. The thesis(Liu et al. 2017; Xu et al. 2018) embedded the attention mechanism(Xu et al. 2015) into the network and let the model decide where to focus its attention. A novel method based on pose-guided spatial attention (PGSA) and activation-based attention (AA) is proposed, which can effectively suppress the occluded region and enhance the significance of the visible region(Xu, Zhao, and Qin 2021).

Methodology

Due to the traditional convolutional neural network requires the same size of input samples, and the partial person re-identification dataset cannot meet this requirement, so we removed the fully connected layer in the convolutional neural network. On this basis, the proposed focus shared area model (FSA) of this thesis is constructed, which mainly includes region pre-definition, foreground-aware region feature extraction, similarity measurement and training strategy, as showed in Figure 2. In Section 3.1, the whole image is divided into a fixed number of regions by uniform segmentation, and pseudo labels are assigned to each region by self-supervised learning. In Section 3.2, the location of visible region and feature extraction of foreground region are introduced. The similarity is introduced in Section 3.3. Our training strategy is introduced in section 3.4.

Region pre-defined by self – supervised learning

Inspired by the PCB(Sun et al. 2018), a region can be considered a component as long as it is stable enough. We divide a fixed number of regions horizontally on the whole image, and assign pseudo labels to each

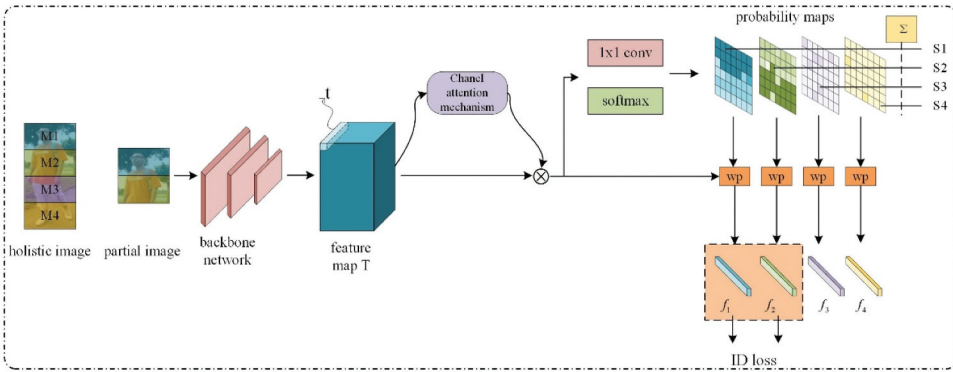


Figure 2. The framework of our proposed FSA. Firstly, the holistic image is pre-defined into a fixed number of regions, and the feature map T is generated through the backbone network. Secondly, the foreground feature map X is generated by the channel attention mechanism, and each region is sensed through 1×1 convolution layer and softmax function, and the distribution probability map and corresponding visibility score of each region is output. Finally, the feature map X is weighted with the distribution probability maps to generate a fixed number of regional features.

pre-defined region through self-supervised learning (Noroozi and Favaro 2016; Wang, He, and Gupta 2017; Wang et al. 2020). The pseudo labels are used as the classification supervision signal, so that the network can easily distinguish the visible regions. Self-supervised learning is a special kind of unsupervised learning method, which can automatically generate supervised signals for feature learning by exploring visual information. Specifically, each region on the input image is projected to the corresponding position on the feature map X through ROI. Assuming that the upper left corner and lower right corner of the region are located at (u_1, v_1) and (u_2, v_2) respectively, the positions on the corresponding feature map X are $(\lceil \frac{u_1}{C} \rceil, \lceil \frac{v_1}{C} \rceil)$ and $(\lceil \frac{u_2}{C} \rceil, \lceil \frac{v_2}{C} \rceil)$ respectively, where C is the lower sampling rate. Then, each pixel t in X is assigned a pseudo label L to represent the region that t belongs to, and Z is used to represent the collection of visible regions. Self-supervised learning plays a crucial role adopted in this thesis. It can not only assign labels to each pre-defined region, but also makes the model focus on the visible region and the shared visible region in the subsequent training of classification loss and triplet loss.

Foreground-aware region feature representation

With the hope of eliminating the interference from background and the non-shared region and pedestrian image spatial misalignment, we employ a channel attention mechanism to pay more attention to perceiving the pedestrian prospects. Firstly, we use ResNet101 as the backbone to extract the feature map T from the input image I , and then obtain the foreground

feature map X through foreground perception. Secondly, we use a 1×1 convolution layer and a softmax function to classify and predict the region of each pixel t on the feature map X :

$$P(M_i|t) = \text{soft max}(W^T t) = \frac{\exp W_i^T t}{\sum_{j=1}^K \exp W_j^T t} \quad (1)$$

Where $P(M_i|t)$ is the predicted probability that t belongs to M_i , W is the weight matrix of the 1×1 convolution layer, K is the number of pre-defined regions. By sliding each pixel t on feature map X , the corresponding probability of X belonging to each pre-defined region can be predicted, and K probability map are obtained, as shown in [Figure 2](#). The visibility score of each region is predicted by accumulating all the values on T , as shown in Equation (2).

$$S_i = \sum_{t \in T} P(M_i|t) \quad (2)$$

If there are a significant number of pixels in this region, we think it is likely to be visible on the input image and the visibility score will be relatively high. On the other hand, if a region is actually invisible, then all the values on the corresponding probability map will be approximately 0, and the visibility score will also be small.

Multiplying the predicted probability map of each region and the feature map obtained by global average pooling to generate the corresponding fine-grained features of each region by the following formula 3.

$$f_i = \frac{\sum_{t \in T} P(M_i|t)t}{S_i}, i = \{1, 2, \dots, K\} \quad (3)$$

where S_i is used to ensure that the number of generated region features is consistent with the number of predefined regions. Even if some regions are actually missing, the corresponding region features will be output in the end, just as the probability map is generated. However, the subsequent similarity measurement is not accepted, as will be explained in [Section 3.3](#).

Similarity measurement

We measure the similarity by matching the common visible region of the query image and the gallery image, and calculated the Euclidean distance of them. The similarity measurement is illustrated in [Figure 3](#). The features of each region i of the two comparison images obtained by our model are f_i^q, f_i^g and the corresponding visibility score are S_i^q, S_i^g .

The local distance D_i^{qg} between each region i of the query image and the gallery image is calculated as follows

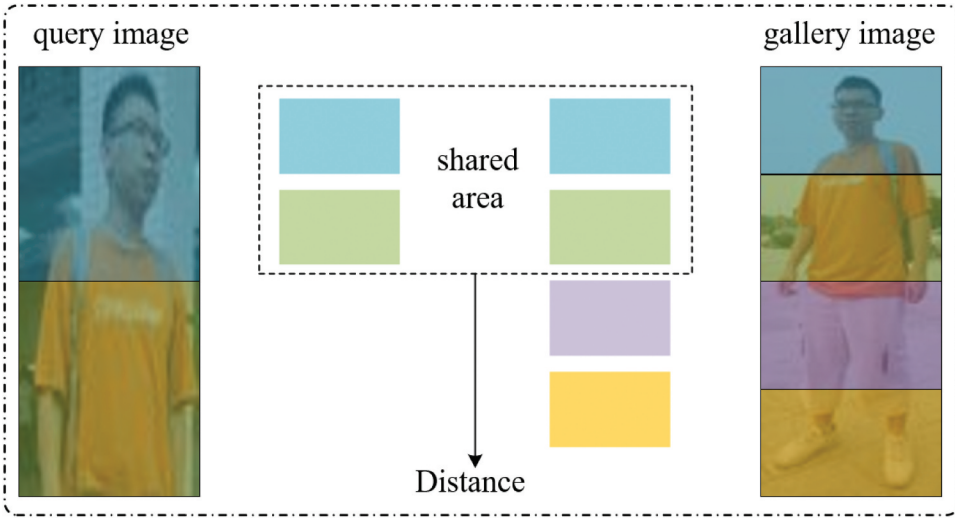


Figure 3. The similarity measure process of query image and gallery image. The similarity measurement is done by focusing on the shared visible region between the query image and the gallery image, and the non-shared region is not accepted at this stage.

$$D_i^{qg} = \|f_i^q - f_i^g\|_2 \quad (4)$$

The finally similarity between the query image and the gallery image is evaluated by calculating D_{qg} .

$$D_{qg} = \frac{\sum_{i=1}^K S_i^q S_i^g D_i^{qg}}{\sum_{i=1}^K S_i^q S_i^g} \quad (5)$$

If the region is not visible, it makes little contribution to the overall distance. Therefore, the overall distance between the images obtained in the end is mainly dominated by the common region.

Training strategy

We adopt a training strategy that uses both cross-entropy loss and triplet loss to train the network model jointly. The cross-entropy loss is utilized to calculate the classification loss of each visible regional feature and the each pixel, which is defined as L_{id} and L_{pixel} :

$$L_{partial}^{id} = CE(l, y) = - \sum_{i \in Z} \Gamma_{l=y} \log(\text{soft max}(P(f_i))) \quad (6)$$

$$L_{pixel} = - \sum_{t \in T} \Gamma_{i=L} \log(P(M_i|t)) \quad (7)$$

where CE denotes the cross-entropy loss, l is the predicted ID of input image and y represent the ground-truth. Z is the set of all visible region. L is the region pseudo label.

In addition, in order to better distinguish similar but different types of input, we adopts triplet loss. Assuming that the three input images are I_a, I_p, I_n , where I_a, I_p are a pair of positive sample pairs and I_a, I_n is a pair of negative sample pairs. It is worth noting that the triplet loss is defined by calculating the distance of the common region between two images as following formula (8):

$$L_{triplet} = [D_{ap} - D_{an} + \delta]_+ \quad (8)$$

$$D_{ap} = \frac{\sum_{i \in |Z^a \cap Z^p|} D_i^{ap}}{|Z^a \cap Z^p|} \quad (9)$$

$$D_{an} = \frac{\sum_{i \in |Z^a \cap Z^n|} D_i^{an}}{|Z^a \cap Z^n|} \quad (10)$$

where D_{ap} is the distance of the common region between the positive sample pair and D_{an} is the distance of the common region between the negative sample pair. Z is the set of visible region. D_i is the distance of each region. After learning through triplet loss, the distance between positive sample pairs is closer, while the distance between negative sample pairs is farther. It is easy to know that the calculation of formula (9) and formula (10) is similar to formula (5). The difference is that the distance calculation in triplet loss is guided by labels during training, and the similarity measurement is guided by visibility scores during testing.

The final loss function is as follows:

$$L = L_{partial}^{id} + L_{pixel} + L_{triplet} \quad (11)$$

Experiments

Datasets and metrics

In this thesis, we verify the effectiveness of our proposed method on the two large-scale public holistic datasets named Market1501, DukeMTMC-reID and two important partial datasets called Partial-REID and Partial-iLIDS, respectively. Moreover, we employ two kinds of evaluation metrics including the cumulative matching characteristics (CMC) and mean average precision (mAP). In our experiment, we only use a single query for image retrieval.

Market1501: Market1501 dataset (L.Zheng et al. 2015a) is collected by six cameras in the campus of Tsinghua University. The dataset contains a total of 32,668 images and provides the training set and the test set. There are 751

people in the training set and 750 people in the test set. The training set contains 12936 images and the test set contains 19732 images. The image is automatically detected and cut by the detector.

DukeMTMC-reID: DukeMTMC-reID (Z.Zheng, Zheng, and Yang 2017) dataset come from 8 different cameras in Duke University. The bounding box of pedestrian images are manually labeled. This dataset consist of training set and test sets. The training set contains 16522 images, and the test set contains 17661 image. There are a total of 702 people in the training data. The dataset provides annotations for pedestrian attributes (gender, backpack, etc).

Partial-REID: The Partial-REID dataset (W.Zheng et al. 2015b) was collected at Sun Yat-Sen University. It included 300 pedestrian images and a total of 30 pedestrians. On average, each pedestrian was manually cropped by five obstructions and five overall images. Due to occluded parts are not the same, the part of the body area obtained is also different. All partial images are query images, and full-body images are test images. Since the dataset is too small, there is no training subset.

Partial-Ilids: The Partial-iLIDS dataset (W.Zheng, Gong, and Xiang 2011) was collected at foreign airports, covering a total of 119 pedestrians and 238 pictures. Each pedestrian has a partial picture and a whole picture. The special feature of this dataset is that most of the occlusions are suitcases carried by pedestrians, so pedestrians in the dataset are mainly the upper body area. Like the Partial-REID dataset, all partial images are used for searching, and all whole images are used for testing. Moreover, in the experiment, generally only the CMC index is used to evaluate the two partial person re-identification datasets. In order to reduce the random error generated by the experiment, it is also necessary to take the average of 10 tests.

Experiment settings

All of our implementations are based on a deep learning framework – PyTorch. We choose to build a focused sharing model based on ResNet101 (He et al. 2016) pre-trained on ImageNet. Because the scale of the partial pedestrian dataset is too small for network training, we take the pictures in the overall pedestrian dataset Martket1501 and DukeMTMC-reID according to a certain proportion R to randomly crop the partial pedestrian images for training. The value range is 0.6–1.0. In our experiment, we set the batch size of each iteration to 64 and use stochastic gradient descent (SGD) to optimize the model. The basic learning rate to 0.1, which attenuated to 0.01 after 30 epochs and ended the training at 140 epochs. We set the momentum and weight decay factors to 0.9 and 0.0005, respectively. We used two loss functions, cross-entropy loss and triplet loss to train the network. The evaluation indicators we use are Rank-1, Rank-5, Rank-10, and mAP. All experiments are performed on an NVIDIA TITAN X GPU.

Comparison with state-of-the-art

In order to train the FSA model we proposed, the whole image is cropped with different proportions R , and partial pedestrian images are obtained as input samples. Where R is set as the value within the interval of 0.6–1. When $R = 1$, it means that the image has not been cropped and the sample is complete. When $R = 0.9$, it means that the image accounts for 90% of the original image, and so on, different degrees of partial person re-identification can be simulated. The smaller R is, the smaller the size of partial pedestrian image is, and the more difficult it is for partial person re-identification.

Market1501 for holistic person Re-ID: As can be seen from Table 1, the comparison results between the method we proposed and previous methods on Market1501 dataset for holistic person Re-ID. There are three comparison methods: baseline model of holistic person Re-ID, PCB model(Sun et al. 2018) based on local feature for holistic person Re-ID, and the more advanced VPM (Sun et al. 2019) for partial Re-ID. Our method achieves $mAP = 83.5\%$ and $Rank-1 = 94.8\%$, while the PCB are $mAP = 83.0\%$ and $Rank-1 = 94.4\%$. Overall, their performance is similar, indicating that our method is also applicable to the holistic person re-identification.

Market1501 for partial person Re-ID: Under different cropping ratios, the sample images are partially missing, which can be regarded as partial person Re-ID. In addition, the smaller the cropping ratio is, the more parts of the image are cropped, and the less information it contains. As shown in Table 2, when more areas of the input sample are cropped, our method will also become worse in mAP and $Rank-1$. However, when compared with the other three methods in horizontal comparison, our method has basically improved in case of different R value. It is worth noting that when the cropping ratio is 0.6, our method achieves 65.1% mAP and 80.1% $Rank-1$ accuracy. Compared with the competitive method VPM, our method is better than it on mAP , but worse than it on $Rank$ accuracy. After analysis, this may be caused by the distribution of Martket1501 dataset. The excellent results in other cases indicate that the method we proposed can alleviate the challenges faced by the partial person re-identification task and is helpful to the partial person re-identification.

Table 1. Comparison on the uncropped Market1501dataset.

Methods	R	Market1501			mAP
		Rank-1	Rank-5	Rank-10	
Baseline	1.0	86.8	95.3	97.4	67.7
PCB+triloss		93.4	97.8	98.4	83.0
VPM		93.0	97.8	98.8	80.0
Ours(FSA)		94.8	98.2	98.8	83.5

Result on DukeMTMC-reID: Just like Market1501, the same large DukeMTMC-reID dataset is clipped to different degrees, and the clipping ratio is also the value within the interval of $[0.6,1]$. When the task is holistic person re-identification, it can be seen from Table 3, our method achieves 86.1% and 74.6% in mAP and Rank -1 , respectively, which is also improved compared with PCB model. For the partial person re-identification task, the robustness of the method was tested with different degrees of region deletion. The experimental results are shown in Table 4. Compared with VPM, the method we proposed basically has certain performance improvement on mAP and Rank-1.

Partial-REID&Partial-iLIDS: It can be clearly seen from Tables 5 and 6 that our method is compared with the existing partial person re-identification methods, including MTRC (Liao et al. 2013), AMC+SWM (Zheng et al. 2015b), DSR (He et al. 2018a), SFR (He et al. 2018b) and VPM (Sun et al. 2019). Our method achieves Rank-1 = 73.7% and Rank-3 = 82.7% on Partial Re-ID dataset. The experiment results on Partial-iLIDS are Rank-1 = 68.9% and Rank-3 = 82.4%, which fully prove the superiority of our method and the performance of partial person re-identification has improved.

Table 2. Comparison on Market1501 dataset cropped in different proportions.

Methods	R	Market1501			mAP
		Rank-1	Rank-5	Rank-10	
Baseline	0.6	79.0	91.4	94.3	57.9
PCB+triloss		8.1	16.5	23.2	6.6
VPM		84.4	94.3	96.1	62.5
Ours(FSA)		80.1	93.0	95.8	65.1
Baseline	0.7	83.9	93.9	95.9	63.7
PCB+triloss		36.8	58.9	67.4	26.8
VPM		88.2	95.8	97.2	71.7
Ours(FSA)		89.1	96.2	97.8	74.3
Baseline	0.8	85.7	94.3	96.4	66.1
PCB+triloss		71.9	87.3	91.4	56.8
VPM		90.1	95.8	97.7	74.7
Ours(FSA)		92.8	97.4	98.5	78.4
Baseline	0.9	87.1	95.5	97.4	67.7
PCB+triloss		88.8	95.8	97.1	77.2
VPM		91.7	96.6	98.0	78.7
Ours(FSA)		93.8	97.9	98.7	81.0

Table 3. Comparison on the uncropped DukeMTMC-reID.

Methods	R	DukeMTMC-reID			mAP
		Rank-1	Rank-5	Rank-10	
Baseline	1.0	76.2	87.3	91.2	58.6
PCB+triloss		84.1	92.4	94.5	73.2
VPM		83.6	91.7	94.2	72.6
Ours(FSA)		86.1	93.4	95.2	74.6

Table 4. Comparison on DukeMTMC-reID dataset cropped in different proportions.

Methods	R	DukeMTMC-reID			mAP
		Rank-1	Rank-5	Rank-10	
Baseline	0.6	76.2	87.3	90.4	55.4
PCB+triloss		13.1	25.6	33.5	10.5
VPM		78.2	89.0	91.3	60.9
Ours(FSA)	0.7	78.3	89.6	92.2	63.9
Baseline		76.3	87.3	90.6	56.7
PCB+triloss		35.9	57.0	65.4	28.4
VPM	0.8	80.3	89.5	92.0	63.1
Ours(FSA)		82.6	91.0	94.0	68.5
Baseline		76.3	88.3	91.9	58.8
PCB+triloss	0.9	64.0	82.6	87.7	52.3
VPM		80.3	89.3	92.4	63.5
Ours(FSA)		84.2	92.1	94.4	71.2
Baseline	0.9	77.0	88.1	91.7	59.0
PCB+triloss		81.6	90.4	93.0	70.3
VPM		81.7	90.9	93.1	70.7
Ours(FSA)		86.7	93.5	95.4	74.1

Table 5. Comparison on Partial-REID.

Methods	Partial-REID	
	Rank-1	Rank-3
MTRC	23.1	27.3
AMC+SWM	37.3	46.0
DSR	50.7	70.0
SFR	56.9	78.5
VPM	67.7	81.9
Ours(FSA)	73.7	82.7

Ablation analysis

We have adopted a method of uniform image segmentation to pre-define the area of the whole person image. In order to explore the influence of the number of areas on the accuracy of the final re-identification task, a series of ablation experiments were conducted on the Market1501 dataset and the DukeMTMC-reID dataset. [Table 7](#) and [Table 8](#) show the results of the ablation analysis, respectively. K refers to the number of areas and its value are 2, 4, 6, 8.

The number of partitioned affects the performance of the network model. When K is 2, no matter on the Rank-1 or mAP, the experimental effect is obviously poor. When K is 4 and 8, the experimental result is also pretty.

Table 6. Comparison on partial-iLIDS.

Methods	Partial-iLIDS	
	Rank-1	Rank-3
MTRC	17.7	26.1
AMC+SWM	21.0	32.8
DSR	58.8	67.2
SFR	64.9	74.8
VPM	67.2	76.5
Ours(FSA)	68.9	82.4

Table 7. Experimental results of different K values on Market1501.

Methods	Market1501	
	Rank-1	mAP
Ours(K = 2)	89.5	73.0
Ours(K = 4)	94.2	83.1
Ours(K = 6)	94.8	83.5
Ours(K = 8)	94.4	82.9

Table 8. Experimental results of different K values on DukeMTMC-reID.

Methods	DukeMTMC-reID	
	Rank-1	mAP
Ours(K = 2)	82.8	66.6
Ours(K = 4)	85.6	74.1
Ours(K = 6)	86.1	74.6
Ours(K = 8)	85.1	73.8

However, considering the experimental effect and cost, we finally choose a compromise scheme, which pre-defined six regions on the entire image as the model choice.

The channel attention mechanism allows the model to focus on the foreground area of the target pedestrian. Considering whether the channel attention mechanism (CAM) has promoted our method, we conducted ablation experiments on the market1501 data set. Table 9 shows the results of the ablation analysis, and the best results are bold in the table.

Table 9. Ablation studies on channel attention mechanism(Market1501).

Methods	R	Market1501			
		Rank-1	Rank-5	Rank-10	mAP
FSA(w/o CAM)	0.6	80.5	92.8	95.6	65.7
FSA(w/ CAM)		80.8	92.9	95.3	65.7
FSA(w/o CAM)	0.7	89.0	95.8	97.4	74.5
FSA(w/ CAM)		89.5	96.5	97.8	74.6
FSA(w/o CAM)	0.8	92.3	97.3	98.7	79.3
FSA(w/ CAM)		92.7	97.1	98.4	79.3
FSA(w/o CAM)	0.9	94.3	97.9	98.8	82.0
FSA(w/ CAM)		93.9	97.9	98.9	82.1
FSA(w/o CAM)	1.0	94.2	98.1	98.9	83.2
FSA(w/ CAM)		94.9	98.2	99.0	83.9

Table 10. Comparison on the uncropped CUHK03 dataset(Detected).

Methods	R	CUHK03	
		Rank-1	mAP
TRiNet(Hermans, Beyer, and Leibe 2017)+RE	1.0	61.8	57.6
PCB		63.7	57.5
Dare(Wang et al. 2018c)+RE		63.3	59.0
Ours(FSA)		64.1	62.6

‘w/o CAM’ means that a CAM module has been removed and ‘w/ CAM’ means that a CAM module has been added. It can be seen that CAM plays a positive role in this dataset, when $R = 1.0$, results of mAP of FSA method with CAM improve 0.7%.

Result on CUHK03 dataset(Detected): The experimental results are shown in Table 10, where RE(Zhong et al. 2017) means random erasing data augmentation.

Conclusion

The goal of partial pedestrian re-recognition is to accurately match the overall image of the pedestrian and the partial image of the pedestrian, but it causes two major problems: spatial misalignment and non-shared area interference. In order to solve these two problems, we focused our attention on the areas shared by the overall image and the partial images. The matching between the shared areas not only does not require alignment but also avoids the interference caused by non-shared areas. Our method mainly includes two parts: region pre-defined and foreground-aware region feature extraction. Firstly, the image is segmented into a fixed number of regions by uniform segmentation, and pseudo-labels are assigned to each region by self-supervised learning, and pseudo-labels are used as supervised signals to guide regional feature learning; Secondly, we use a channel attention mechanism to allow the model to focus on the target pedestrian foreground area, the probability that each pixel belong to each pre-defined region was predicted, and the probability maps and visibility scores were obtained. The features of each area are weighted by pedestrian foreground features and probability maps. Experiments demonstrate the effectiveness and superiority of the proposed method on holistic pedestrian datasets: Market1501, DukeMTMC-reID, and two partial pedestrian datasets: Partial-REID and Partial-iLIDS.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 61972056, in part by the Hunan Provincial Natural Science Foundation of China under Grant 2021JJ30743 and 2021JJ30741, in part by the Degree & Post-graduate Education Reform Project of Hunan Province of China under Grant 2020JGZD043.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Natural Science Foundation of Hunan Province [2021JJ30743 and 2021JJ30741]; the Degree & Post-graduate Education Reform Project of Hunan Province of China [2020]GZD043]; the National Natural Science Foundation of China [61972056].

ORCID

Shuren Zhou  <http://orcid.org/0000-0002-0465-3258>

References

- Ge, Y., Z. Li, H. Zhao, G. Yin, S. Yi, X. Wang, and H. Li. 2018. FD-GAN: Pose-guided feature distilling GAN for robust person re-identification. *ArXiv*: 1810.02936.
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–78.
- He, L., J. Liang, H. Li, and Z. Sun. 2018a. Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7073–82.
- He, L., Z. Sun, Y. Zhu, and Y. Wang. 2018b. Recognizing partial biometric patterns. *ArXiv*: 1810.07399.
- Hermans, A., L. Beyer, and B. Leibe. 2017. In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737*.
- Hu, J., L. Shen, and G. Sun. 2018. Squeeze-and-excitation networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 7132–41.
- Iodice, S., and K. Mikolajczyk. 2018. Partial person re-identification with alignment and hallucination. *Asian Conference on Computer Vision*, Perth Australia, pp. 101–16.
- Ji, Z., X. Zou, X. Lin, X. Lin, T. Huang, and S. Wu. 2020. An Attention-driven Two-stage Clustering Method for Unsupervised Person Re-Identification. *Proceedings of the European Conference on Computer Vision*, Online, pp. 20–36.
- Kalayeh, M. M., E. Basaran, M. Gökmen, M. E. Kamasak, and M. Shah. 2018. Human semantic parsing for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 1062–71.
- Kumar, V., A. Namboodiri, M. Paluri, and C. V. Jawahar. 2017. Pose-aware person recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 6223–32.
- Li, D., X. Chen, Z. Zhang, and K. Huang. 2017. Learning deep context-aware features over body and latent parts for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 384–93.
- Li, W., X. Zhu, and S. Gong. 2018. Harmonious attention network for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 2285–94.
- Liao, S., A. K. Jain, and S. Z. Li. 2013. Partial face recognition: Alignment-free approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (5):1193–205. doi:10.1109/tpami.2012.191.

- Lin, Y., L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang. 2019. Improving person re-identification by attribute and identity learning. *Pattern Recognition* 95:151–61. doi:10.1016/j.patcog.2019.06.006.
- Liu, C., X. Chang, and Y.-D. Shen. 2020. Unity style transfer for person re-identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 6887–96.
- Liu, J., B. Ni, Y. Yan, P. Zhou, S. Cheng, and J. Hu. 2018. Pose transferrable person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 4099–108.
- Liu, X., H. Zhao, M. Tian, L. Sheng, J. Shao, S. Yi, J. Yan, and X. Wang. 2017. Hydraplus-net: Attentive deep features for pedestrian analysis. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 350–59.
- Miao, J., Y. Wu, P. Liu, Y. Ding, and Y. Yang. 2019. Pose-guided feature alignment for occluded person re-identification. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, South Korea, pp. 542–51.
- Miao, J., Y. Wu, and Y. Yang. 2021. Identifying visible parts via pose estimation for occluded person re-identification. *IEEE Transactions on Neural Networks and Learning Systems* 99: 1–11. doi:10.1109/TNNLS.2021.3059515.
- Noroozi, M., and P. Favaro. 2016. Unsupervised learning of visual representations by solving jigsaw puzzles. *Proceedings of the European Conference on Computer Vision*, Amsterdam, Netherlands, pp. 69–84.
- Qian, X., Y. Fu, T. Xiang, W. Wang, J. Qiu, Y. Wu, Y.-G. Jiang, and X. Xue. 2018. Pose-normalized image generation for person re-identification. *Proceedings of the European Conference on Computer Vision*, Munich, Germany, pp. 661–78.
- Sarfraz, M. S., A. Schumann, A. Eberle, and R. Stiefelhagen. 2018. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 420–29.
- Si, J., H. Zhang, C.-G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang. 2018. Dual attention matching network for context-aware feature sequence based person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 5363–72.
- Song, C., Y. Huang, W. Ouyang, and L. Wang. 2018. Mask-guided contrastive attention model for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 1179–88.
- Song, S., W. Zhang, J. Liu, and T. Mei. 2019. Unsupervised person image generation with semantic parsing transformation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 2357–66.
- Sun, Y., L. Zheng, Y. Yang, Q. Tian, and S. Wang. 2018. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). *Proceedings of the European Conference on Computer Vision*, Munich, Germany, pp. 501–18.
- Sun, Y., Q. Xu, Y. Li, C. Zhang, Y. Li, S. Wang, and J. Sun. 2019. Perceive where to focus: Learning visibility-aware part-level features for partial person re-identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 393–402.
- Wang, D., and S. Zhang. 2020. Unsupervised person re-identification via multi-label classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 10981–90.

- Wang, G., Y. Yuan, X. Chen, J. Li, and X. Zhou. 2018a. Learning discriminative features with multiple granularities for person re-identification. *Proceedings of the 26th ACM international conference on Multimedia*, Seoul, South Korea, pp. 274–82.
- Wang, X., K. He, and A. Gupta. 2017. Transitive invariance for self-supervised visual representation learning. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 1329–38.
- Wang, Y., L. Wang, Y. You, X. Zou, V. Chen, S. Li, G. Huang, B. Hariharan, and K. Q. Weinberger. 2018c. Resource aware person re-identification across multiple resolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 8042–51.
- Wang, Y., Z. Chen, F. Wu, and G. Wang. 2018b. Person re-identification with cascaded pairwise convolutions. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 1470–78.
- Wang, Z., J. Zhang, L. Zheng, Y. Liu, Y. Sun, Y. Li, and S. Wang. 2020. CycAs: Self-supervised Cycle Association for Learning Re-identifiable Descriptions. *Proceedings of the European Conference on Computer Vision*, Online, pp. 72–88.
- Wei, L., S. Zhang, H. Yao, W. Gao, and Q. Tian. 2017. Glad: Global-local-alignment descriptor for pedestrian retrieval. *Proceedings of the 25th ACM international conference on Multimedia*, California, USA; pp. 420–28.
- Woo, S., J. Park, J.-Y. Lee, and I. S. Kweon. 2018. CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision*, Munich, Germany, pp. 3–19.
- Xu, J., R. Zhao, F. Zhu, H. Wang, and W. Ouyang. 2018. Attention-aware compositional network for person re-identification. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 2119–28.
- Xu, K., J. Ba, R. Kiros, K. Cho, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio. 2015. Show, attend and tell: Show, attend and tell: Neural image caption generation with visual attention. *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, Lille, France, pp. 2048–57.
- Xu, Y., L. Zhao, and F. Qin. 2021. Dual attention-based method for occluded person re-identification. *Knowledge-Based Systems* 212:106554. doi:10.1016/j.knsys.2020.106554.
- Yan, C., G. Pang, J. Jiao, X. Bai, X. Feng, and C. Shen. 2021. Occluded person re-identification with single-scale global representations. *Proceedings of the IEEE International Conference on Computer Vision*, Montreal, Canada, pp. 11875–84.
- Yu, H.-X., and W.-S. Zheng. 2020. Weakly supervised discriminative feature learning with state information for person identification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, USA, pp. 5528–38.
- Zhang, X., H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun. 2017. AlignedReID: Surpassing human-level performance in person re-identification. *ArXiv*: 1711.08184.
- Zhao, H., M. Tian, S. Sun, J. Shao, J. Yan, S. Yi, X. Wang, and X. Tang. 2017a. Spindle net: Person re-identification with human body region guided feature decomposition and fusion. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 1077–85.
- Zhao, L., X. Li, Y. Zhuang, and J. Wang. 2017b. Deeply-learned part-aligned representations for person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 3219–28.
- Zheng, F., C. Deng, X. Sun, X. Jiang, X. Guo, Z. Yu, F. Huang, and R. Ji. 2019a. Pyramidal person re-identification via multi-loss dynamic training. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 8514–22.

- Zheng, L., L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian. 2015a. Scalable person re-identification: A benchmark. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 1116–24.
- Zheng, L., Y. Huang, H. Lu, and Y. Yang. 2019b. Pose-invariant embedding for deep person re-identification. *IEEE Transactions on Image Processing* 28 (9):4500–09. doi:10.1109/tip.2019.2910414.
- Zheng, L., Y. Yang, and A. G. Hauptmann. 2016. Person re-identification: Past, present and future. *ArXiv*: 1610.02984.
- Zheng, W.-S., S. Gong, and T. Xiang. 2011. Person re-identification by probabilistic relative distance comparison. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, USA, pp. 649–56.
- Zheng, W., X. Li, T. Xiang, S. Liao, J. Lai, and S. Gong. 2015b. Partial person re-identification. *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, pp. 4678–86.
- Zheng, Z., L. Zheng, and Y. Yang. 2017. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. *Proceedings of the IEEE International Conference on Computer Vision*, Venice, Italy, pp. 3754–62.
- Zheng, Z., L. Zheng, and Y. Yang. 2018. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology* 29 (10):3037–45. doi:10.1109/tcsvt.2018.2873599.
- Zhong, Z., L. Zheng, G. Kang, S. Li, and Y. Yang. 2017. Random erasing data augmentation. arXiv preprint arXiv:1708.04896.
- Zhu, K., H. Guo, Z. Liu, M. Tang, and J. Wang. 2020. Identity-guided human semantic parsing for person re-identification. *Proceedings of the European Conference on Computer Vision*, Online, pp. 346–63.
- Zhu, Z., T. Huang, B. Shi, M. Yu, B. Wang, and X. Bai. 2019. Progressive pose attention transfer for person image generation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, USA, pp. 2347–56.