MULTI-LEVEL FEATURE FUSION BASED TRANSFER LEARNING FOR PERSON RE-IDENTIFICATION

Yingzhi Chen and Tianqi Yang

Department of Computer Science, Jinan University, China

ABSTRACT

Most of the currently known methods treat person re-identification task as classification problem and used commonly neural networks. However, these methods used only high-level convolutional feature or to express the feature representation of pedestrians. Moreover, the current data sets for person re-identification is relatively small. Under the limitation of the number of training set, deep convolutional networks are difficult to train adequately. Therefore, it is very worthwhile to introduce auxiliary data sets to help training. In order to solve this problem, this paper propose a novel method of deep transfer learning, and combines the comparison model with the classification model and multi-level fusion of the convolution features on the basis of transfer learning. In a multi-layers convolutional network, the characteristics of each layer of network are the dimensionality reduction of the previous layer of results, but the information of multi-level features is not only inclusive, but also has certain complementarity. We can using the information gap of different layers of convolutional neural networks to extract a better feature expression. Finally, the algorithm proposed in this paper is fully tested on four data sets (VIPeR, CUHK01, GRID and PRID450S). The obtained re-identification results prove the effectiveness of the algorithm.

KEYWORDS

Person re-identification; transfer learning; multi-level feature fusion & deep learning

1. INTRODUCTION

Person re-identification is an important area of research in computer vision. The goal of person re-identification is to identify the same person from the pictures of candidates, and the probe and gallery are capture different cameras. It is worth noting that the shooting areas of these cameras often do not intersect. In recent years, person re-identification systems have become more widely used in criminal investigation and law enforcement agencies. The complete system consists of person detection, person tracking and person retrieval. Each part is complex, and the three parts of the computer vision field have evolved independent tasks. The person re-identification task content studied in this paper is roughly the same as the person retrieval task. It consists of two steps: feature extraction and distance matching between candidate sets. In the first step, there are generally two difficulties of person re-identification: 1) Detail features such as face, fingerprint and iris are not stable due to the resolution and angle of view of person image. 2) Person pictures are collected by cameras with different perspectives. When person are captured at different positions, the body posture and angle will be different. Person remain unchanged under different cameras in terms of overall characteristics such as clothing, shape and skin. Thus, more distinguishing features are nesseary. For the distance matching step, given a target picture and a candidate picture set, the correct picture in the candidate set should be closest to the target picture in the feature space. The distance measure of the picture can be supervised or unsupervised. This article focuses on supervised learning. The basic process of pedestrian recognition is shown in Figure 1.

DOI: 10.5121/ijaia.2019.10302

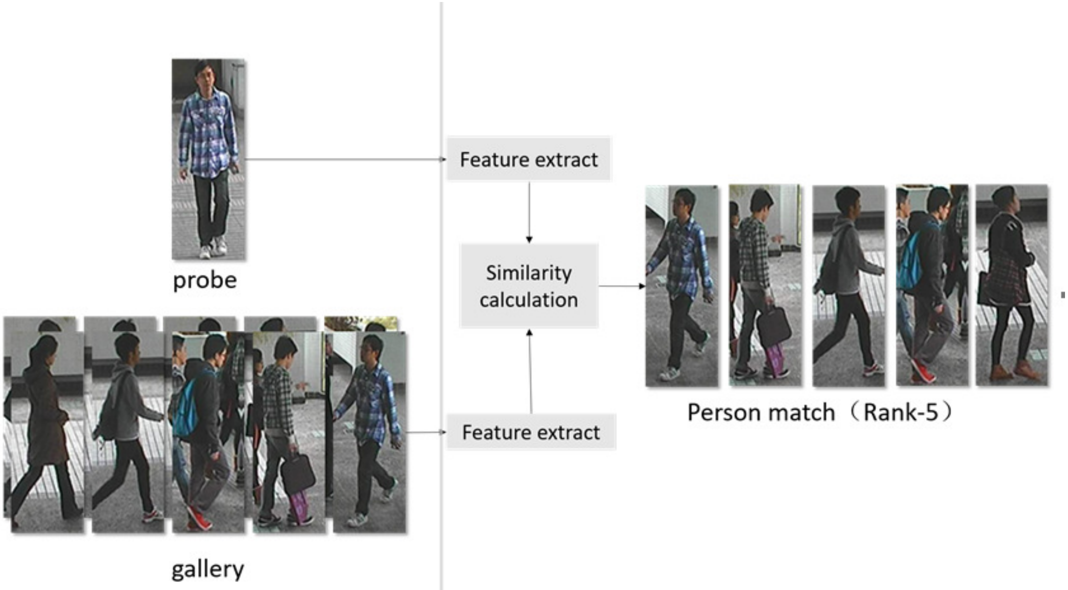

Figure1. The process of person re-identification

Since 2012, Krizhevsky et al. [1] used the convolutional network to win the championship in the ILSVRC (ImageNet Large Scale Visual Recognition Competition)-2012 image classification competition. Deep learning has become one of the research hotspots in many field [24, 25, 26]. Deep learning method has a good effect. At present, person re-identification task also begins to adopt the deep learning method, and has achieved good experimental results. Li et al. [2] first proposed a FPNN (Filter pairing neural network) model based on convolutional network, which first uses convolution and pooling operation turns the picture into a series of local features, and then the image is divided into multiple regions for matching, which used a unified neural network framework to handle the effects of background noise, lighting, and occlusion. Ahmed et al. [3] proposed an enhanced neural network framework that takes advantage of the proximity differences in input picture pairs and picture blocks to simplify features. Compared with manual feature extraction, the process of deep learning to extract features is more direct and convenient, and it is no longer necessary to manually design features based on person images. The characteristics of deep learning can be automatically adapted to the influence of angle, illumination and occlusion in person re-identification. This is because deep learning has a large number of learnable parameters. By continuously adjusting parameters, a target feature related to a specific target task can be obtained. Despite the advantages of deep learning, it relies on the number of samples, especially the number of samples in the classification task, and the small-scale data is difficult to train a suitable neural network. Transfer learning is different from general machine learning, is aims to learn knowledge form one domain and use the knowledge on other different domain. Transfer learning had wide applications in computer vision [21, 22, 23]. The training data and test data of machine learning are generally from the same task, and the training data of the transfer learning and the source of the test data are inconsistent, there is a large gap between the training data and the target data. How to reduce this distribution difference is the key concern of transfer learning. When applying transfer learning in a neural network, there are usually two ways to reduce the distribution difference. One is to select the auxiliary data as close as possible to the target data, and the other is to perform secondary training on the target data, that is, fine-tuning. The transfer learning process in deep learning has two steps: pre-training and fine-tuning. First, the task of learning in the past is called the source task, the data is the source data, the task to prepare for learning is called the target task, and the data is the target data. Pre-training means that the model is trained on the source data in advance to complete the source task. Fine-tuning refers to the second training of the pre-trained model on the target data to complete the target task. From the perspective of initialization, pre-training is equivalent to having the model parameters have an initial value that is more in line with the target task, and the fine-tuning process is training on this initialization. In the case where the target data and source data are more consistently distributed, there are two main advantages of fine-tuning: (1) Save time costs and use pre-trained models without reinitializing the network for training. (2)With better generalization ability, the pre-trained model is generally trained on large data sets. Combined with the data set of the target task, it is equivalent to using additional auxiliary data to complete the target task.

At present, the mainstream method of using a multi-layer convolution network is to convolve the feature map layer by layer from the input picture, and then input the final layer convolution result as a final feature of the pedestrian picture to the classifier. Most of gestures, clothing and accessories of person are very similar, using the most prominent features of the unity does not make the most of the full use of person information. This article considers this issue, so in this chapter, we explore the multi-level features of convolutional networks. Convolutional networks use only three simple operations: convolution, pooling, and full connection. Convolution and pooling alternate to form the main part of the network. These operations perform a large number of complex operations, extracting abstract features, making the interior of the product network is more like a "black box." In recent years, many researchers have used convolution visualization to explain the mechanism of convolutional networks. A lot of work has been done on the features learned by convolutional networks. Convolution visualization can be traced back to AlexNet [1] proposed by Krizhevshy et al., which directly visualizes the convolution kernel of the first convolutional layer of the network, and displays the value of the convolution kernel as a picture. The response strength of the product is visually visible, but the smaller convolution kernel of this method, such as 3×3 and 5×1, is difficult to apply because the resolution of the small convolution kernel after conversion to a picture is very low, and it is difficult to obtain meaningful results. The work of Grishick et al. [4] then marked the area of the AlexNet that responded strongly to the input data. Zeiler et al. [5] used the same convolution kernel transposition as the network itself as a new convolution operation, taking the convolved feature map as input, de-convolve the same form of image as the original input, and improving the visualization based on the visualization results of AlexNet, and achieved good results in related tasks. We using multi-level feature fusion of convolutional neural network and twice transfer learning to learn more information of person.

2. OUR METHOD

2.1. Two-Step Transfer Learning

Figure 2 shows the two-step transfer learning process. It is divided into three phases. The specific settings are shown in Table 1. After the auxiliary data set is determined and the training model obtains a basic model, the basic model needs to be fine-tuned on the target data to obtain the final model. The whole process is divided into two phases: the pre-training phase and the fine-tuning phase. In the pre-training phase, our neural network model is consistent with the models used in other general visual classifications. The data set is not specially processed. The network structure used the ResNet [6] structure and the loss function is the classification loss. The architecture of ResNet is shown in Table 1.

Table 1. The architecture of ResNet

Layer	Kernel	Channel	Repeat	Pooling
CNN	7×7	64	1	3x3, max
CNN	1×1 3×3 1×1	64 64 256	3	none
CNN	1×1 3×3 1×1	128 128 512	4	none
CNN	1×1 3×3 1×1	256 256 1024	6	none
CNN	1×1 3×3 1×1	512 512 2048	3	none
Fully-connect	none	none	1	average

In the fine-tuning stage, for person re-identification, it can be divided into two models to fine-tune: classification model and verification model. Because the test data set is too small during the one-step transfer process and spans three types of data sets, the simple classification model and the verification model can't achieve good results, so this paper proposed a new joint model combining the two models. Thus the general process of two-step deep transfer learning is divided into pre-training, fine-tuning and feature extraction. Since the tasks for pre-training are different from the tasks for fine-tuning and feature extraction, the network parameters and structure of deep learning will retain most of them and change a small part, so that pre-training learning can retain knowledge, while fine-tuning the network can learn new knowledge, the changed part is often the last layer of the full connection, because the last layer of the fully connected layer of neuron network in the classification task and the number of species and the last layer of information strong correlation with task. In this paper, the classification model is used in the pre-training phase. In the fine-tuning phase, the same number of neurons as the pedestrian re-identification identity is used to form the full-connection layer instead of the original layer. The activation function is softmax. After the desired category is output, the intersection is calculated. And the output of the last layer of convolution is used as the input of the above new full connection, and the second is the characteristic input of the contrast loss, thereby calculating the cross entropy while calculating the contrast loss, and finally adding the two losses as the final loss, using this loss to fine-tune the pre-trained network. In the last phase, we extract the feature maps of convolutional layers. The hyper-parameters of pre-training same as [6]. And in fine-tuning, the hyper-parameters are shown in Table 2.

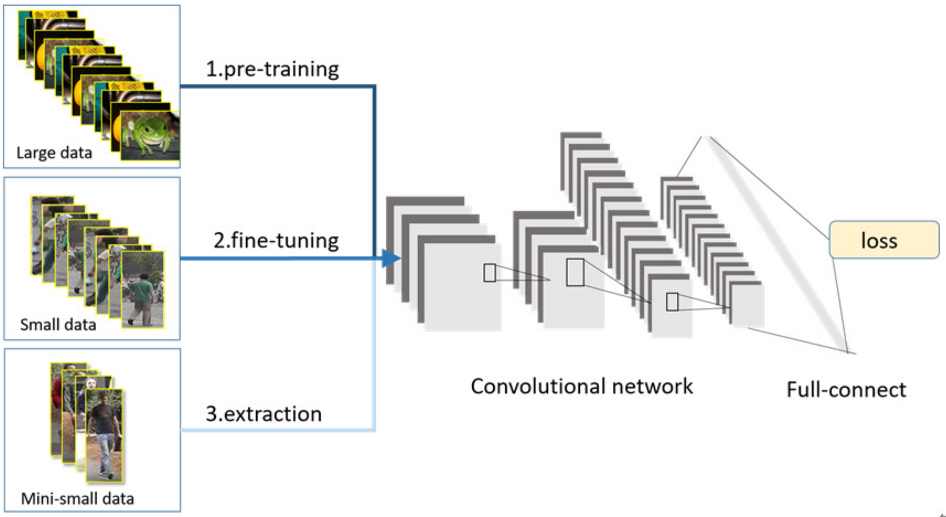

Figure 2. Two-step Transfer Learning

Table 2. The Hyper-Parameters in Fine-Tuning

Learning rate	Decay of weight	Decay of learning rate	Optimization	Batch
0.0003	0.0005	0.1	Adam[27]	32

2.1. Multi-level Feature Fusion

The convolutional network has the characteristics of parameter sharing and local sensing. In addition, it can be found that the characteristics of the multi-layer convolution network have the following special properties:

(1) On the image, the convolution is calculated by weighting and summing a pixel and its adjacent pixels. The intuitive interpretation of this method is to change the value of this pixel according to the information around the pixel. If the value of the convolution kernel is 0 except for the target pixel being 1, then the information of each pixel output is the same as the input information; if the value of the convolution kernel is different, the output pixel will contain information of other pixels, assuming the convolution kernel value is [1,0,1], then for each pixel of the input, the output will be the sum of the pixel values on both sides of the pixel, that is, the information of the pixel is filtered out and replaced by the pixel information on both sides, so the volume The product feature is the filtering result of the input of this layer. After learning, it will filter the noise that is meaningless to the task result, which can be regarded as the abstraction of the input data.

(2) In a network consisting of a full-convolutional, the output characteristics of each layer are the result of convolution operations on the input of this layer, and the convolution has an abstraction effect, also assuming that the value of each layer of convolution kernel is [1,0,1], the pixel value of the first layer convolution result is the sum of the pixel values on both sides, the new pixel value filters out the original middle pixel, and the second layer convolution result is the filtering of the first layer result, nth The layer output is the filtered result of the n-1th layer output, and this abstraction exists in each layer of convolution. As the number of network layers grows deeper, the convolution feature has more and more abstract features, that is, it has progressive abstract features.

(3) Under ideal conditions, the multi-level convolution network is well trained and the optimal solution is obtained. Then the last layer is characterized by all the information that affects the recognition effect. However, the training of neural network is a non-convex optimization problem. The actual situation often cannot reach the ideal state. Therefore, convolution filtering the input data noise information will also filter out some meaningful information, which will affect the recognition effect. Multi-level features will have some complementarity. It can be summarized as a multi-level feature with three characteristics: abstraction, progressiveness and complementarity.

Given an input x, the output of the convolutional layer is:

$$x^n = f(W^n x^{n-1} + b^n), n \in (1, N) \tag{1}$$

Where f denotes the activation function. After the fine-tuning of the multi-layer convolution network, the traditional approach is to input x into the network, and then extract the output of the last layer of convolution as x^{N} a feature to represent x, we have some complementarity according to multi-level features. We extract the output of the multi-layer convolution and fuse it into X as the feature of the person picture in the form of splicing, the form of X is:

$$\mathrm{X} = [x^1, x^2, \dots, x^N] \tag{2}$$

That is:

$$\mathrm{X} = [f(W^1 x^0 + b^1), f(W^2 x^1 + b^2), \dots, f(W^N x^{N-1} + b^N)] \tag{3}$$

The input image is a colour image, that is, there are three colour channels of R, G, and B. In each layer, the number of channels according to different feature patterns of the channel parameters changes accordingly, and each feature map size of the random multi-level feature is different, and each feature map is doubled by a pooling layer with a step size of 2×2. This section uses the pooled feature map as a feature representation, so that the characteristics of each layer are due to the dimension. The proportion of each layer in the fusion feature is different. Because the degree of influence of each layer on the recognition result is unknown, this section extracts each layer after the feature is extracted, so that the feature dimensions of each layer after pooling are consistent. Use $pool^n$ to represent the nth layer of pooling, then:

$$\mathrm{X} = [pool^1(f(W^1 x^0 + b^1)), pool^2(f(W^2 x^1 + b^2)), \dots, pool^N(f(W^N x^{N-1} + b^N))] \tag{4}$$

The multi-level fusion method is shown in Figure 3. After the person image input into the network, the output of the selected pooling layer is extracted. This output is originally flowed as an input to the next layer network. This algorithm uses it as a method. Part of the expression of the overall characteristics of pedestrians, the progressiveness proposed in the above summary is slowly progressive in multi-layer convolution networks, and the degree of increase in the degree of abstraction between each adjacent convolutional layer is not very large. That is, the closer the information is to the convolution feature on the network structure, the more similar the information is, smaller the information difference is. There are information gap in each convolutional layers.

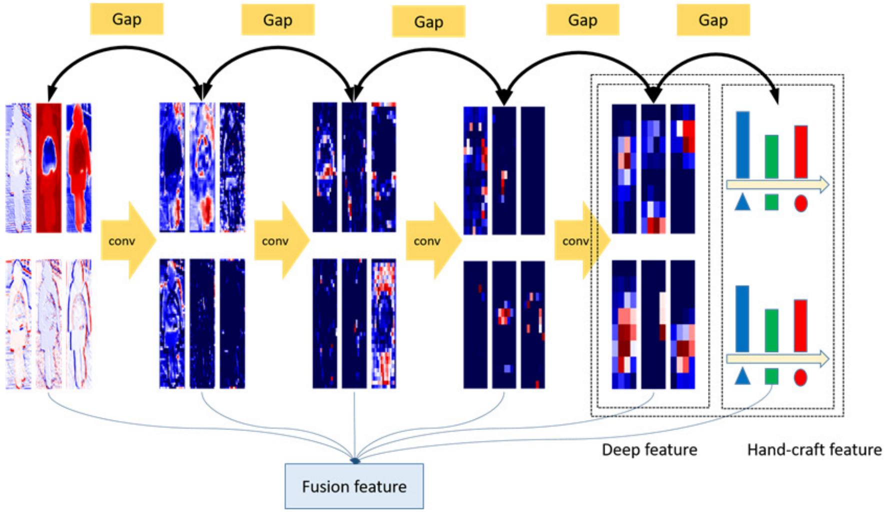

Figure 3. Feature Fusion

3. EXPERIMENTS

3.1. Datasets

Experiments are conducted in the following three datasets:

VIPeR: The VIPeR dataset [7] was collected from the outdoor environment by the University of California Computer Vision Lab and contained 632 pedestrians under two cameras. The same pedestrian consists of two pictures. The size of all the images is 128x48, and the angles, postures and illumination of the two pictures of the same pedestrian are not the same. Most of the two pictures have a viewing angle of about 90 degrees, among which pedestrians More will carry a backpack and the appearance of the backpack changes at different angles.

CUHK01: The CUHK01 data set [8] is collected from two cameras on the campus of the Chinese University of Hong Kong. It consists of a total of 971 pedestrians, and each pedestrian has two pictures under the same camera, and each pedestrian is photographed by two cameras, that is, each pedestrian contains 4 pictures with a total of 3884 images, because it is shot on campus. The captured pedestrians carry more school bags.

PRID450S: The PRID450S dataset [9] is based on the PRID2011 dataset and contains 450 pairs of pedestrian images for a total of 900 images. All images are captured by two fixed cameras with no cross-field of view. The original captured data is in video format. The image data is obtained by cropping. The resolution of the cropped images is not consistent and the shooting location is on an outdoor street. Most pedestrians pose. It is the side, and because the shooting time has a certain interval, the pedestrian will move and cause the pedestrian background to be different.

GRID: The GRID data set [10] is a smaller data set that is collected from eight disjoint cameras at the underground station. It contains 250 pedestrian image pairs, that is, each pedestrian has two images captured from different cameras, and also contains 775 additional images, which do not belong to 250 pedestrians, and therefore exist as interference. GRID has a large change in pedestrian attitude, color and illumination, and the resolution of the picture is low. The underground parking lot is dense, so the picture contains more than one pedestrian. Other pedestrians are used as the background, and there are cases where they are blocked by other pedestrians.

3.2. Experimental results

This summary compares the multi-level fusion feature on the four data and the deep and handcraft fusion feature with the mainstream method from 15 to 18 years, as shown in Table 3. In the case of fine-tuning, in terms of Rank-1 scores, the deep and handcraft features achieved the best results on the VIPeR, CUHK01, and PRID450S datasets, with the best results compared to other publicly available results increased by 9.67%, 3.9%, 4.7%, respectively.. The Rank-1 score achieved the second best result on the GRID dataset. The multi-level fusion feature on VIPeR and CUHK01 also achieved the best score compared with other public results. The GRID dataset contains a large proportion of unidentified information. Person pictures lead to a more specific identification. Although this method does not achieve the optimal level on Rank-1, it has achieved a good results. The comparison methods selected in this paper include deep learning methods and manual feature methods. Although the former results are often better than the latter, the latter will still achieve better results in some tests. Under the non-fine tuning setting, the fusion feature of this paper has achieved the best results on the VIPeR, CUHK01 and PRID450S datasets compared to the mainstream methods in the past three years. The non-fine tuning result on CUHK01 is higher than the fine-tuning result of 1.6% on PRID450S. The ratio is 0.1% higher than the fine-tuning result, and the other two data sets are lower than the fine-tuning result. This verifies the multi-level feature fusion and the effectiveness of the manual-depth feature fusion, and the improvement effect of the joint model on migration learning.

Table 3. Overall Performance

		VIPeR			CUHK01			PRID450S			GRID		
Method		Rank-1	Rank-10	Rank-20	Rank-1	Rank-10	Rank-20	Rank-1	Rank-10	Rank-20	Rank-1	Rank-10	Rank-20
LOMO+XQDA[11]	ICCV15	40.00	80.51	91.08	63.21	-	-	61.4	91.0	95.3	18.96	52.56	62.24
KHPCA[12]	ICPR16	39.4	85.1	93.5	-	-	-	52.2	92.8	94.4	-	-	-
LSSCDL[13]	CVPR16	42.7	84.3	91.9	65.97	-	-	60.5	88.6	93.6	22.4	51.3	61.2
GOG+XQDA[14]	CVPR16	49.7	88.7	94.5	67.3	91.8	95.9	68.4	94.5	97.8	24.7	58.4	69.0
DNS[15]	CVPR16	51.17	90.51	95.92	69.09	91.77	95.39	-	-	-	-	-	-
DLPAR[16]	ICCV17	48.7	85.1	93.0	75.0	95.7	97.7	-	-	-	-	-	-
JLML[17]	IJCAI17	50.2	84.3	91.6	76.7	95.6	98.1	-	-	-	37.50	69.40	77.40
SSM[18]	CVPR17	53.73	91.49	96.08	72.98	96.76	99.11	-	-	-	27.20	61.12	70.56
EDPRMLS[19]	CVPR18	50.10	84.35	-	-	-	-	-	-	-	-	-	-

GCT[20]	AAAI18	49.4	87.2	94.0	61.9	87.6	92.8	58.4	84.3	89.8	-	-	-
Ours(Deep)		59.1	91.1	96.1	77.9	95.6	98.0	59.0	88.8	93.9	20.8	49.1	58.1
Ours(Fusion & Fine Tune)		63.4	93.3	97.1	80.6	96.6	98.5	73.1	95.0	98.3	27.4	57.6	69.0
Ours(Fusion & No-Fine Tune)		57.6	92.2	96.8	82.2	97.6	99.1	73.2	96.0	99.0	25.4	57.4	69.8

4. CONCLUSIONS

In order to apply transfer learning more effectively, this paper proposes a joint model based on the comparison model and classification model, and applies the joint model to the fine-tuning phase of migration learning. Therefore, the model can learn the category-related features as well as the features that can make the same class features closer in the feature space and the different class features are farther away. This paper visualizes and discusses the results of different layers of multi-layer convolutional networks, and summarizes the characteristics of multi-level features with abstraction, progressiveness and certain complementarity. According to the characteristics of certain complementarity, a method of merging multi-level features is proposed, which expands the feature expression of convolutional networks. This paper selects four widely used public datasets, namely VIPeR, GRID, PRID450S and CUHK01, to verify the algorithm comprehensively. The experimental results show that the joint model and feature fusion have obvious performance improvement compared with single feature. By comparing the algorithm results published by researchers in the field of pedestrian recognition in recent years, it is proved that the method has good recognition accuracy.

REFERENCES

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105).

[2] Li, W., Zhao, R., Xiao, T., & Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 152-159).

[3] Ahmed, E., Jones, M., & Marks, T. K. (2015). An improved deep learning architecture for person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3908-3916).

[4] Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 580-587).

[5] Zeiler, M. D., & Fergus, R. (2014, September). Visualizing and understanding convolutional networks. In European conference on computer vision (pp. 818-833). springer, Cham.

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).

[7] Gray, D., & Tao, H. (2008, October). Viewpoint invariant pedestrian recognition with an ensemble of localized features. In European conference on computer vision (pp. 262-275). Springer, Berlin, Heidelberg.

[8] Li, W., Zhao, R., & Wang, X. (2012, November). Human reidentification with transferred metric learning. In Asian conference on computer vision (pp. 31-44). Springer, Berlin, Heidelberg.

[9] Roth, P. M., Hirzer, M., Köstinger, M., Beleznai, C., & Bischof, H. (2014). Mahalanobis distance learning for person re-identification. In Person re-identification (pp. 247-267). Springer, London.

[10] Van De Weijer, J., Schmid, C., Verbeek, J., & Larlus, D. (2009). Learning color names for real-world applications. IEEE Transactions on Image Processing, 18(7), 1512-1523.

[11] Liao, S., Hu, Y., Zhu, X., & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2197-2206).

[12] Prates, R. F., & Schwartz, W. R. (2016, December). Kernel hierarchical PCA for person re-identification. In 2016 23rd International Conference on Pattern Recognition (ICPR) (pp. 2091-2096). IEEE.

[13] Zhang, Y., Li, B., Lu, H., Irie, A., & Ruan, X. (2016). Sample-specific svm learning for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1278-1287).

[14] Matsukawa, T., Okabe, T., Suzuki, E., & Sato, Y. (2016). Hierarchical gaussian descriptor for person re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1363-1372).

[15] Zhang, L., Xiang, T., & Gong, S. (2016). Learning a discriminative null space for person re-identification. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1239-1248).

[16] Zhao, L., Li, X., Zhuang, Y., & Wang, J. (2017). Deeply-learned part-aligned representations for person re-identification. In Proceedings of the IEEE International Conference on Computer Vision (pp. 3219-3228).

[17] Li, W., Zhu, X., & Gong, S. (2017). Person re-identification by deep joint learning of multi-loss classification. arXiv preprint arXiv:1705.04724.

[18] Bai, S., Bai, X., & Tian, Q. (2017). Scalable person re-identification on supervised smoothed manifold. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2530-2539).

[19] Guo, Y., & Cheung, N. M. (2018). Efficient and deep person re-identification using multi-level similarity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2335-2344).

[20] Zhou, Q., Fan, H., Zheng, S., Su, H., Li, X., Wu, S., & Ling, H. (2018, April). Graph correspondence transfer for person re-identification. In Thirty-Second AAAI Conference on Artificial Intelligence.

[21] Saenko, K., Kulis, B., Fritz, M., & Darrell, T. (2010, September). Adapting visual category models to new domains. In European conference on computer vision (pp. 213-226). Springer, Berlin, Heidelberg.

[22] Gopalan, R., Li, R., & Chellappa, R. (2011, November). Domain adaptation for object recognition: An unsupervised approach. In 2011 international conference on computer vision (pp. 999-1006). IEEE.

[23] Hoffman, J., Guadarrama, S., Tzeng, E. S., Hu, R., Donahue, J., Girshick, R., ... & Saenko, K. (2014). LSDA: Large scale detection through adaptation. In Advances in Neural Information Processing Systems (pp. 3536-3544).

[24] Penha. A & Penha. D. (2018, March). Home Appliance Identification for Nilm Systems Based on Deep Neural Networks. International Journal of Artificial Intelligence & Applications (IJAIA), (pp. 69-80).

[25] Salahat. S & Awad. M. (2017, March). Short-Term Forecasting of Electricity Consumption in Palestine Using Artificial Neural Networks. International Journal of Artificial Intelligence & Applications (IJAIA), (pp. 11-21).

[26] Artificial Neural Networks. International Journal of Artificial Intelligence & Applications (IJAIA), (pp. 17-29).

[27] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

INTENTIONAL BLANK