An Efficient Person Re-Identification Method Based on Deep Transfer Learning Techniques

Shimaa Saber
Information Technology Department
Faulty of Computers and Information
Menoufia, Egypt
shimaa.saber@ci.menofia.edu.eg

Khalid Amin
Information Technology Department
Faulty of Computers and Information
Menoufia, Egypt
kh.amin.0.0@gmail.com

Mohamed Hammad
Information Technology Department
Faulty of Computers and Information
Menoufia, Egypt
mohammed.adel@ci.menofia.edu.eg

Abstract—**Person re-identification (re-id) is a significant process in applications of video analysis. Several applications in different areas such as airports and stations are used multiple cameras in different places for monitoring and investigation, which are expensive and can be easily abused. Therefore, automatic person re-identification techniques are highly required. The main issue of this field is to find distinguishing features that represent the person. In this paper, we proposed an efficient method to extract the main features based on the deep transfer learning technique for a person re-id system. In addition, we employed a support vector classifier (SVC) as a separated classifier for the final decision to increase the accuracy of the system. We employed several publicly available datasets, which are the main datasets used for person re-id purposes in the literature. The proposed method achieved the best accuracy of 89.59% for rank-1, which outperforms the state-of the-art methods. Finally, the simulation results reveal that the proposed system is efficient prior to person re-id.**

Keywords— person re-identification; transfer learning; SVC; deep learning; video analysis.

I. INTRODUCTION

Recently, the research in computer vision has seen impressive developments in the techniques of pattern recognition, tracking and machine learning [1- 3]. This growth in computer vision research led to the production of more effective applications and systems for intelligent surveillance systems and forensic industries, and as a result, the demand for these products increased. Several applications in different areas such as airports, department stores, public parks and stations are used multiple cameras in different places for monitoring and investigation, which are expensive and can be easily abused [2]. Therefore, continuous recognition, human tracking, and person re-identification techniques are highly required. In this study, we focused on person re-identification (re-id).

Person re-id technique is the process of comparing an image of a person with other images obtained from different angles and cameras, then choosing the same person. The main issue of this field is to find distinguishing features that represent the person. Several recent approaches used conventional-based algorithms for extracting the features and obtaining a good accuracy. These methods used to extract and encode feature descriptor for personal appearance, most of this feature works on low-level appearance and reduce the distance between a feature of the same person, and this distance is done by machine learning algorithms that are the classification process [4]. However, these approaches confront several challenges such as working on big data and overfitting problems. Deep learning approaches can solve these previous

problems and can achieve state-of-the-art accuracy. Also, images that capture from these cameras, face several challenging problems like low-resolution, occlusions, pose variations and viewpoint change as shown in Fig. 1.

Fig. 1. some challenging problem in VIPeR dataset. (A) low resolution image, (B) Illumination variations, (C) Occlusions, (D) changing in Pose and viewpoint

Certain most of the state-of-art techniques are presented based on the most common deep learning approach called convolutional neural network (CNN), due to its powerful feature learning abilities and fitting capacity. However, several methods using a complex design of their deep model, which increases the cost and time of the process. In this paper, we overcome most of the previous problems by employing a pre-trained deep model with transfer learning for extracting deep features and achieving good performance. The pre-trained model is a network that has already learned to extract powerful and informative features from the input images, which makes our system much faster and easier than other models that trained from scratch. The main contributions of this work are as follows:

- We propose an efficient system that employed a pre-trained deep CNN model for human re-id, which overcomes the problems that faced most of the previous traditional machine learning methods and achieve state-of-art accuracy in person re-id.

- We present a deep transfer learning technique for person re-id by employing VGG-Net as a person descriptor by selecting the appreciated layers to obtain a good representation of the data.
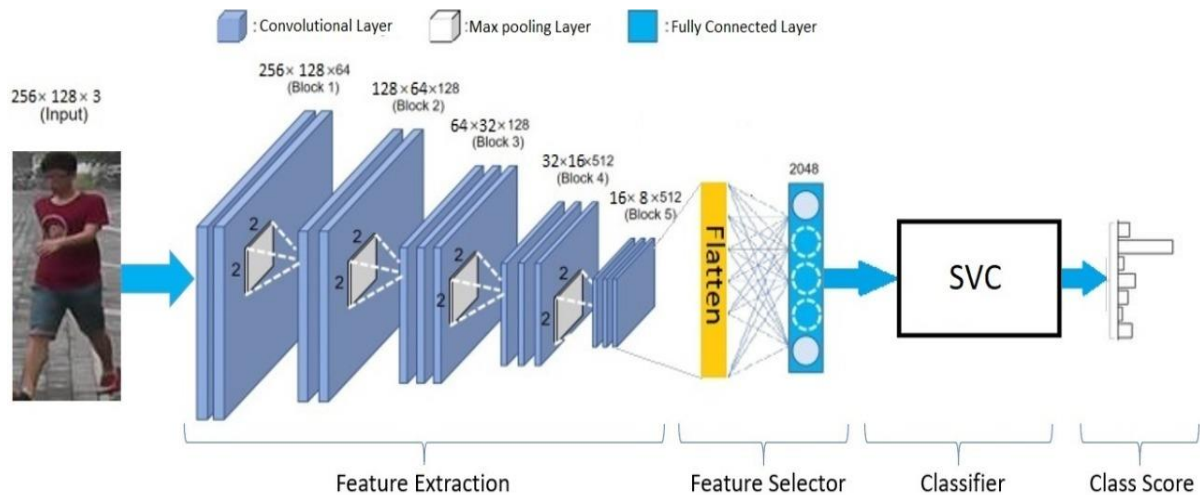
Fig. 2. Block diagram of the proposed pre-identification system

- We employ a support vector classifier (SVC), which can solve the problem of using small sample data and augment the overall accuracy of the proposed system.

- We evaluate our approach on three different datasets CUHK01 [5], i-LIDS [7] and Market-1501 [8], which are the main datasets used in person re-id purposes in the literature. Results show that the proposed approach outperforms other approaches in terms of accuracy for rank-1.

This paper is organized as follows: Section II provides a discussion of the previous methods. Section III provides a discussion of the proposed method. In Section IV, we report and analyze the experimental results. Finally, conclusions are drawn in Section V.

II. RELATED WORK

Recently, with the development of artificial intelligence, personal re-id has become a research focus and a variety of classification methods for this topic have been developed. The re-id methods used to extract and encode feature descriptor for person appearance [9]. This method is divided into conventional-based methods and deep learning-based methods. Conventional person re-id methods mainly depend on the features of manual design and measurement learning, most of this feature extractor works on low-level appearance such as color, texture, and shape [10]. The measurement learning reduce distance between feature of the same person such as support vector machine (SVM), neural network (NN), K-nearest neighbor (KNN) and etc.. Most existing conventional methods focus on extracting the features manually and cannot capture advanced semantic information and not robust to body poses, variances of viewpoints and illumination, also distance metric learning suffers from computational complexity and poor generalization ability. The deep learning-based methods harness learning technology that is used for training the models to recover images and re-identify persons in an end-to-end system [4]. The deep learning model, person re-id methods are divided into three parts: convolution neural based methods, person re-id based on CNN and recursive neural network (RNN), generative adversarial network (GAN)-based methods and the hybrid of deep learning and traditional method [9]. Deep learning has strong abilities of independent learning and feature extraction

but lack the structure to extract temporal features, it requires very large amount of data to perform better than other techniques and is extremely expensive to train due to complex data models. In this Section, we discussed several recent works based on the two classification methods as follows:

A. Conventional-based methods

- Yang *et al*. [10] presented a method for describing colors in a person's image using salient color names based color descriptor (SCNCD). This method calculates the color histograms from different color spaces to make the final features strong to variation in illumination. Feature extraction in SCNCD is done in several steps to generate the vector of feature, which is the combination of the histogram for the six parts. For matching the person features, SCNCD uses KISSME technique [3]. This method effected by the background influence and variation of illumination. They worked on Viper [11] and PRID [12] and obtained accuracies of 37.8% and 41.6%, respectively for rank-1.

- Liao *et al*. [13] introduced an approach that combines efficient feature extraction and the learning method. They presented two methods as Local Maximal Occurrence (LOMO) and Cross-view Quadratic Discriminant Analysis (XQDA). LOMO is used to extract the features, and XQDA is utilized for a person matching. LOMO has a disadvantage that its robustness to illumination is not excellent. They worked on CUHK03 [6] and obtained accuracies of 46.25% and 52.2% for detected and labeled respectively for rank-1.

- Zhao *et al*. [14] presented an unsupervised method for person re-id. This method extracts different features includes LAB color histogram and SIFT descriptor. These features are extracted from patches of size 10×10 pixels. In the training process, the method takes the extracted feature without requesting the person label. This method does not work well with illumination variation , its computations cost time and gives low performance. They obtained an accuracy of 76.36% when working on rank-20.

B. Deep learning-based methods

- Li *et al*. [15] designed a model to capture the local context knowledge by using a neural network. This model called a Multi-Scale Context Aware Network (MSCAN) that extracted powerful features from the full body and body parts by stacking each layer in the network with multi-scale convolutions. Spatial Transformer Networks (STN) with novel spatial constraints are used to localize and learn pedestrian for deformable parts in body-parts representations. However, the limitations of this study is pose changing of person that is sometimes dangerous in feature matching. They obtained an accuracy of 80.31% for rank-1 when worked on Market1501 [8] dataset.

- Kim *et al*. [16], presented a graph convolution network (GCN) for person re-id. The GCN method utilizes the information from the anchor, negative, and positive samples. They worked on the Market-1501 [8] dataset and obtained an accuracy of 88.65% for rank-1. This study is not robust to the noise, complex and obtained low performance on small data.

- Yang and Chen [18], presented a re-id method based on fusing features extracted from CNN with traditional hand-crafted features. In this method the computational is expense, time is complexity and low accuracy on small data. They obtained an accuracy of 57% for rank-1 when worked on CUHK03 [6] dataset.

- Wang and Bi [19], presented a person re-id method based on graph relation learning. Then, the weight operation has been utilized to get the relation vertices to receive suggestions from other details and finally, these vertices are trained for learning more features. They worked on CUHK03-NP-L and CUHK03-NP-D and obtained accuracies of 75.2% and 71.4%, respectively for rank-1.

III. METHODOLOGY

In this Section, we discussed the databases used in this study and each step of the proposed system in detail. Fig. 2 shows a block diagram of the overall system, which is consists of the following stages: feature extraction stage using the pre-trained deep model to extract the deep features of the input images and classification stage using SVC for final decision. In this study, we worked on images obtained from three different datasets CUHK01 [5], i-LIDS [7] and Market-1501 [8]. We prepared all input images to be suitable for the used deep model by resizing each image to the desired dimension of 256×128.

A. The Used Datasets

In order to develop and evaluate the proposed system, we consider three different public datasets, which are the main datasets used in person re-id purposes in the literature.

- **CUHK01 [5]:** is collected from 2 cameras with disjoint views and contains 971 identities. For each identity, the dataset contains 2 samples extracted from the two different cameras. The quality of the captured image in the dataset is relatively good. The image is scaled to 160 x 60 pixels.

- **i-LIDS [7]:** is an extension from iLIDS MCTS dataset, this dataset contains 300 person. Images in

the dataset are captured by 2 cameras. iLIDS dataset has large occlusion and contains both single shots and multiple shots for a person.

- **Market 1501 [8]:** The videos in the dataset are collected in an open environment at Tsinghua University by 6 dis-joint cameras among which 5 are 1080×1920 HD and 1 is 576×720 SD. This dataset contains 1501 person and the bounding boxes for the person are extracted by both deformable part models (DPM) and manually labeling. This dataset implicates 2793 false images from using the DPM detector, so the quality of the detected images is worse than images in the CUHK03 [6] dataset. Also contains both single shot and multiple shots for a person and the image is normalized to 128 x 64 pixels. Table I describes the three datasets in detail

TABLE I. DESCRIPTION OF THE USED DATASET IN THIS STUDY

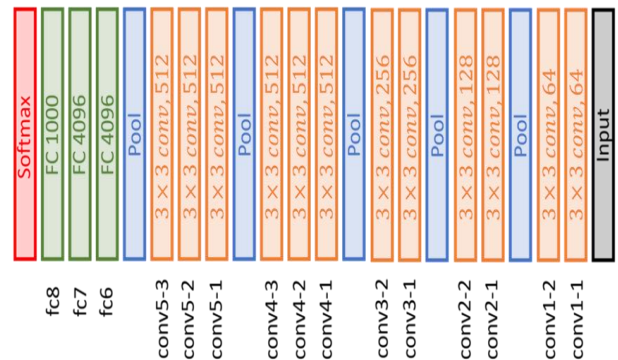| Properties	Datasets		
	CUHK01	*i-LIDS*	*Market 1501*
Year	2012	2014	2015
Camera	2	2	6
Person	971	300	1501
Label	Manually	Manually	Manually /DPM
Images	3884	42795	32217
Image size	160×60	various	160×60
Muti-Shot	✓	✓	✓
Illumination Variations	✓	✓	✓
Pose Variations	✓	✓	✓
Partical Variations	✓	✓	✓
Scale Variations	×	✓	✓

Fig. 3. Architecture of pre-trained VGG-16

B. Extracting the Deep Features

In this stage, the pre-trained deep CNN model referred to as VGG-16 [20] is applied to the re-id system by transfer learning techniques. Comparing with other pre-trained models such as Alexnet [21], a small respective field (3x3 with a stride of 1) is used in VGG-16 instead of a large respective field (11x11 with a stride of 4) as in Alexnet, which helps to improve the overall performance. In addition, based on the previous research, VGG-16 has been proven to have higher accuracy than other pre-trained models.

For our system, we employed the VGG-16 as a feature extraction by removing the output layer in the original model and use the entire network as a fixed feature extractor for the new data set. As shown in Fig. 3 the input of this model is the person images of dimension 256×128 passed through five blocks of convolution (conv) layers with a total of 13 conv layers. The size of the filter that is used in these conv layers is very small (3×3) with the same padding (1-pixel). Each conv block is followed by a max pooling (maxpool) layer of stride (2, 2). Finally, there are three fully connected (fc) layers passed to the Softmax layer for classification and the ReLU activation function was used in all hidden layers. In our case, Softmax layer and the last two fully connected (fc) layers of the pre-trained VGG-16 model are removed in the proposed method and we selected only the first fully connected layers as a feature selector. Here, the output of the first fully connected layer (fc1) gives the best performance comparing with other layers and passes it to the classifier. Finally, we obtained the feature matrix with size $D \times 4096$, where D is the number of input images.

C. Classification

We performed 5-fold cross-validation as well as testing using 80-20 training and validation-testing split. Therefore, the data were divided and pass to the classifier for training. In this stage, a separated classifier called SVC is employed for the classification task. SVC is the support vector machine algorithm, which is employed for the multiclass problem. The choice of loss functions and penalties has more flexibility in this classifier, also it has a better accommodate large numbers of samples. Dense and sparse input is supported in this classifier and one-vs-rest schemes are used to handle the multiclass. We selected the SVC classifier as it performs better for re-id features than other classifiers and does not suffer from overfitting [22]. This is the most commonly used strategy for multiclass classification and is a fair default choice.

IV. EXPERIMENTAL Results AND Discussion

The proposed system is carried out on several datasets including CUHK01 [5], i-LIDS [7] and Market1501 [8]. All input images from the datasets are resized to 256×128. For training the VGG network to extract the feature, the optimizer of the network uses momentum of the stochastic gradient descent (SGD) algorithm to optimize the weight w with momentum 0.9, 0.04 for learning rate and decayed by 0.1. Our network is implemented in the PyTorch framework and trained on Colab that provides NVIDIA Tesla T4. The training loss for the datasets for the first 50 epochs is depicted in Fig. 4-a. Also Fig. 4-b shows classification accuracy performance graphs. The numbers training and testing images for the three datasets are listed in TABLE II. The iteration for training the SVC is 1000, the squared hinge loss for loss function, 1.0 for Penalty of the error term and tolerance for stopping criteria by 1e-4. The training time of SVC for CUHK01 is 8.44 minutes, and 0.026 sec for testing of one image.

A. Evaluation Protocol

The single-shot method is used widely, so in our experiment, we rely on it to allow extensive comparison. The performance of person Re-id algorithms is shown in terms of recognition rate that is accomplished by the Cumulative Matching Characteristic (CMC) [23]. CMC is a common metric utilized to evaluate the recognition rate and shows the percentage of correct matches that are located below a specific rank. The specific rank value is obtained by matching each probe image to the gallery set.

TABLE II. DESCRIPTION OF NUMBER OF TRAINING AND TESTIONG IMAGES

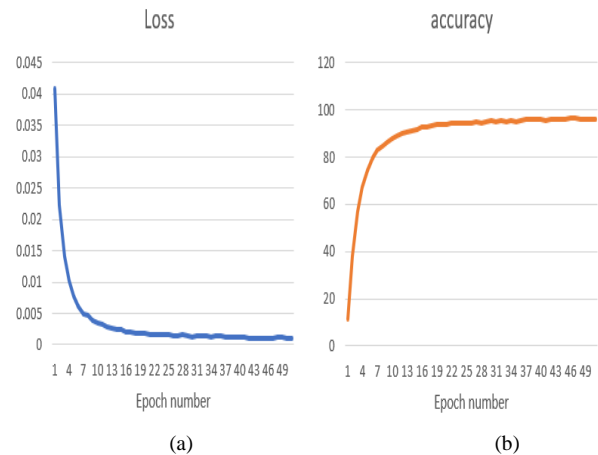| Topic	Datasets		
	CUHK01	i-LIDS	Market 1501
Training	2913	33967	23534
Testing	971	8492	5884

Fig. 4. Training performance graphs: (a) Loss vs iteration graph, (b) accuracy versus iteration graph, during training with SGD

B. Comparing with other methods on the Used Datasets

Comparing to the previous methods, the proposed method appears to have a high performance prior to person re-id. The proposed method is compared with the state-of the- art prior to performance on the three datasets as follows:

CUHK01 database [5]: We compare our proposed method with some of the traditional and deep learning methods, and the competitive results are shown in TABLE III. The proposed method significantly outperforms the performance for rank-1 compared to the previous methods. Our method gives 86.1% in rank-1 matching, which is higher 4.35% than the second method. Notice that our method can give good results even with this smaller dataset.

Market1501 database [8]: TABLE VI provides the competitive accuracy results between the proposed method and the baseline for problems of person re-id using the Market-1501 dataset. As shown in TABLE VI, our method gives an improvement in results compared to the previous methods. Our method gives 89.57% in rank-1 vs 88.65% for GCN [16], also gives better results in other ranks.

i-LIDS database [7]: TABLE VII provides the competitive accuracy results between the proposed method and the baseline for problems of person re-id. As shown in TABLE VII, the proposed method exceeded the previous results significantly, which illuminates the effectiveness of the proposed method. Our method gives 75.92% in rank-1 vs 52.33% for CFL [49], also gives better results in other ranks.

Methods	Rank			
	1	5	10	20
KISSME [3]	14.02	32.2	44.44	56.61
RD+RCCA [24]	31.1	-	68.55	-
SDALF [25]	9.9	22.57	30.33	41.03
eSDC [26]	19.67	32.72	40.29	50.58
LMNN[27]	13.45	31.33	42.25	54.11
ITML [28]	15.98	35.22	45.60	59.81
BOF [30]	16.60	34.0	44.85	56.73
VGG Net [20]	73.92	90.72	95.05	-
DeepReid (FPNN) [31]	27.87	58.2	73.46	-
FFN [32]	55.51	78.40	83.68	92.59
Hybrid [33]	67.12	89.45	91.68	96.54
end-to-end CAN [34]	67.2	87.3	92.5	97.2
MuDeep [35]	79.01	97.0	98.96	-
Appearance model [36]	81.75	94.02	96.49	-
Our(VGG+SVC)	**86.1**	**94.54**	**96.7**	**97.7**

Methods	Rank			
	1	5	10	20
LOMO+XQDA [13]	27.2	41.6	49.1	-
Bow [44]	35.8	52.4	60.3	-
UMDL [45]	34.5	52.6	59.6	-
ECN [37]	75.1	87.6	91.6	94.5
MMFA [38]	56.7	75.0	81.8	-
MAR [39]	67.7	81.9	87.3	-
SPGAN [40]	51.5	70.1	76.8	-
HHL [41]	62.2	78.8	84.0	-
PAUL [42]	68.5	82.4	87.4	-
SVDNet [43]	82.30	92.30	-	-
GCN [16]	88.65	95.72	-	-
Our (VGG+SVC)	**89.57**	**96.98**	**98.12**	**98.93**

V. CONCLUSION

The main objective of this paper is to propose an efficient person re-id method based on the deep transfer learning technique and SVC classifier for classification. We employed the pre-trained VGG-16 to extract the deep features of the input images. After that, the extracted deep features are fed to SVC for the classification. The proposed method is worked on three different datasets CUHK01, i-LIDS and Market-1501, which are the main datasets used for person re-id purposes in the literature. The obtained result shows that the proposed method implemented by VGG-16 and SVC, can help for improving the performance of the person Re-ID system. Based on the obtained results, the proposed method delivers better performance than other previous methods on rank-1 accuracy. We achieved the best accuracy of 89.57 for rank-1 when using Market1501 database. The obtained results show that the proposed method is efficient prior to the person re-id task. In our future work, we try to use other CNN models and observe the performance, work on different datasets for training and testing, and improve our approach.

Methods	Rank			
	1	5	10	20
SDALF [25]	28.80	47.78	56.96	68.04
LMNN [27]	28.0	53.8	66.1	82.3
ITML [28]	29.0	54.0	70.5	86.7
multi-scale CNN [46]	33.63	56.32	69.75	84.22
PRDC [47]	37.8	63.7	75.1	88.4
MCAM [48]	34.0	58.3	67.0	77.0
CFL [49]	53.33	70.00	78.33	88.32
MuDeep [35]	41	70	83	-
Our (VGG+SVC)	**75.92**	**91.40**	**94.86**	**97.19**

REFERENCES

[1] Vezzani, Roberto, D. Baltieri, and R. Cucchiara, "People reidentification in surveillance and forensics: A survey.", ACM Computing Surveys (CSUR), vol. 46 n.2, pp 1-37, 2013.

[2] Kim, Miri, J. Jung, H. Kim, and J. Paik, "Person re-identification using color name descriptor-based sparse representation", 7th Annual Computing and Communication Workshop and Conference (CCWC), pp. 1-4. IEEE, 2017.

[3] M. Köstinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischo,. "Large scale metric learning from equivalence constraints." in: Proceedings of the Conference on Computer Vision and Pattern Recognition, IEEE, pp. 2288-2295, 2012.

[4] Huang, Yan, H. Sheng, Y. Zheng, and Z. Xiong. "DeepDiff: Learning deep difference features on human body parts for person re-identification." Neurocomputing vol. 241, pp. 191-203, 2017.

[5] Li, Wei, R. Zhao, and X. Wang. "Human reidentification with transferred metric learning." In Asian conference on computer vision, Springer, Berlin, Heidelberg, pp. 31-44 , 2011

[6] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in Computer Vision and Pattern Recognition, pp. 152–159, 2014.

[7] Wang, Taiqing, Sh. Gong, X. Zhu, and Sh. Wang, "Person re-identification by video ranking." In European conference on computer vision, Springer, Cham, pp. 688-703, 2014.

[8] Zheng, Liang, L. Shen, L. Tian, Sh. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark." In Proceedings of the IEEE international conference on computer vision, pp. 1116-1124. 2015

[9] Wang, Kejun, H. Wang, M. Liu, X. Xing, and T. Han. "Survey on person re‐identification based on deep learning." CAAI Transactions on Intelligence Technology vol: 3, no: 4, pp: 219-227, 2018.

[10] Yang, Yang, J. Yang, J. Yan, Sh. Liao, D. Yi, and S. Z. Li , "Salient color names for person re-identification." European conference on computer vision, Springer, Cham, 2014.

[11] Gray, Douglas, and H. Tao. "Viewpoint invariant pedestrian recognition with an ensemble of localized features." In European conference on computer vision, Springer, Berlin, Heidelberg, pp. 262-275, 2008.

[12] Hirzer, Martin, C. Beleznai, P. M. Roth, and H. Bischof. "Person re-identification by descriptive and discriminative classification." In Scandinavian conference on Image analysis, Springer, Berlin, Heidelberg, pp. 91-102, 2011.

[13] Liao, Shengcai, Y. Hu, X. Zhu, and S. Z. Li. "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, 2015.

[14] Zhao, Rui, W. Ouyang, and X. Wang. "Unsupervised salience learning for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition, . pp. 3586-3593, 2013.

[15] Khan, Furqan M., and F. Brémond. "Person re-identification for real-world surveillance systems." arXiv preprint arXiv:1607.05975, 20 Jul 2016

[16] Kim, Guisik, D. W. Shu, and J. Kwon. "Robust person re-identification via graph convolution networks." Multimedia Tools and Applications, pp: 1-10, 2021.

[17] Loy, Ch. Change, T. Xiang, and Sh. Gong. "Multi-camera activity correlation analysis." In IEEE Conference on Computer Vision and Pattern Recognition, pp. 1988-1995. IEEE, 2009

[18] Yang, XiuJie, and P. Chen. "Person re-identification based on multi-scale convolutional network." Multimedia Tools and Applications vol. 79, no. 13, pp: 9299-9313, Apr 2020.

[19] Wang, Hao, and Xiaojun Bi. "Person Re-Identification Based on Graph Relation Learning." Neural Processing Letters, vol.53, no. 2, pp: 1401-1415, 2021.

[20] Simonyan, Karen, and A. Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556, 2014.

[21] Krizhevsky, Alex, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems vol. 25, pp: 1097-1105, 2012.

[22] Patro, K. Kumar, S. P. R. Reddi, S. E. Khalelulla, P. R. Kumar, and K. Shankar. "ECG data optimization for biometric human recognition using statistical distributed machine learning algorithm." The Journal of Supercomputing vol.76, no. 2, pp: 858-875, 2020.

[23] Wang, Xiaogang, G. Doretto, Th. Sebastian, J. Rittscher, and P. Tu. "Shape and appearance context modeling." In 2007 ieee 11th international conference on computer vision, pp. 1-8. IEEE, 2007.

[24] An, Le, M. Kafai, S. Yang, and B. Bhanu, "Person reidentification with reference descriptor." IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 4, pp. 776-787, 2015

[25] Bazzani, Loris, M. Cristani, and V. Murino. "Symmetry-driven accumulation of local features for human characterization and re-identification." Computer Vision and Image Understanding, vol. 117 n.2 , pp- 130-144, 2013.

[26] Zhao, Rui, W. Ouyang, and X. Wang. "Unsupervised salience learning for person re-identification." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3586-3593, 2013.

[27] K.Q. Weinberger, L.K. Saul, "Distance metric learning for large margin nearest neighbor classification". Journal of machine learning research, vol. 10 n.2, pp. 207–244, 2009

[28] J.V. Davis, B. Kulis, P. Jain, S. Sra, I.S. Dhillon, "Information-theoretic metric learning", In Proceedings of the 24th international conference on Machine learning, pp. 209–216, 2007

[29] M. Guillaumin, J. Verbeek, and C. Schmid. "Is that you? Metric learning approaches for face identification." In 2009 IEEE 12th international conference on computer vision, pp. 498-505, 2009.

[30] W. Li, R. Zhao, T. Xiao, and X. Wang, "Deepreid: Deep filter pairing neural network for person re-identification," in Computer Vision and Pattern Recognition, pp. 152–159, 2014.

[31] Zhang, Lixia, K. Li, Y. Qi, and F. Wang. "Local feature extracted by the improved bag of features method for person re-identification." Neurocomputing, 2020.

[32] Wu, Shangxuan, Y.C. Chen, X. Li, A. C. Wu, J. J. You, and W. S. Zheng. "An enhanced deep feature representation for person re-identification." In IEEE winter conference on applications of computer vision (WACV), pp. 1-8, 2016

[33] Wu, Lin, Ch. Shen, and A. V. D. Hengel. "Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification." Pattern Recognition, vol. 65, pp. 238-250, 2017.

[34] Liu, Hao, J. Feng, M. Qi, J. Jiang, and Sh. Yan. "End-to-end comparative attention networks for person re-identification." IEEE Transactions on Image Processing, vol. 26, no. 7, pp. 3492-3506, 2017

[35] Qian, Xuelin, Y. Fu, Y. G. Jiang, T. Xiang, and X. Xue. "Multi-scale deep learning architectures for person re-identification." In Proceedings of the IEEE International Conference on Computer Vision, pp. 5399-5408, 2017

[36] Huang, Lei, Wenfeng Zhang, Jie Nie, and Zhiqiang Wei. "Person re-identification based on multi-appearance model." Multimedia Tools and Applications, vol. 80, pp. 16413-16423, 2021.

[37] Zhong, Zhun, L. Zheng, Z. Luo, Sh. Li, and Y. Yang. "Invariance matters: Exemplar memory for domain adaptive person re-identification." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 598-607, 2019.

[38] Lin, Shan, H. Li, Ch. T. Li, and A. Ch. Kot. "Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification." arXiv preprint arXiv:1807.01440, 2018.

[39] Yu, H. Xing, W. Sh. Zheng, A. Wu, X. Guo, Sh. Gong, and J. H. Lai. "Unsupervised person re-identification by soft multilabel learning." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2148-2157. 2019.

[40] Deng, Weijian, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp: 994-1003, 2018.

[41] Zhong, Zhun, L. Zheng, Sh. Li, and Y. Yang. "Generalizing a person retrieval model hetero-and homogeneously." In Proceedings of the European Conference on Computer Vision (ECCV), pp. 172-188, 2018.

[42] Yang, Qize, H. X. Yu, A. Wu, and W. Sh. Zheng. "Patch-based discriminative feature learning for unsupervised person re-identification." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3633-3642, 2019.

[43] Sun, Yifan, L. Zheng, W. Deng, and Sh. Wang. "Svdnet for pedestrian retrieval." In Proceedings of the IEEE international conference on computer vision, pp. 3800-3808, 2017.

[44] Zheng, Liang, L. Shen, L. Tian, Sh. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark." In Proceedings of the IEEE international conference on computer vision, pp. 1116-1124, 2015.

[45] Peng, Peixi, T. Xiang, Y. Wang, M. Pontil, Sh. Gong, T. Huang, and Y. Tian, "Unsupervised cross-dataset transfer learning for person re-identification." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1306-1315, 2016.

[46] X. Yang, and P. Chen. "Person re-identification based on multi-scale convolutional network." Multimedia Tools and Applications 79, no. 13, pp: 9299-9313, 2020.

[47] Ding, Shengyong, L. Lin, G. Wang, and H. Chao. "Deep feature learning with relative distance comparison for person re-identification." Pattern Recognition vol. 48, no. 10, pp: 2993-3003, 2015.

[48] F. M. Khan, and F. Brémond. "Person re-identification for real-world surveillance systems." arXiv preprint arXiv:1607.05975, 2016

[49] Franco, Alexandre, and L. Oliveira. "A coarse-to-fine deep learning for person re-identification." In 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1-7, 2016