Image and Vision Computing 58 (2017) 224–229

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Multi-label convolutional neural network based pedestrian attribute classification☆

Jianqing Zhu[a], Shengcai Liao[b,*], Zhen Lei[b], Stan Z. Li[b]

[a]College of Engineering, Huaqiao University, Quanzhou, Fujian 362021, China
[b]Center for Biometrics and Security Research, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

ARTICLE INFO

Article history:
Received 6 October 2015
Received in revised form 3 June 2016
Accepted 16 July 2016
Available online 25 July 2016

Keywords:
Pedestrian attribute classification
Multi-label classification
Convolutional neural network

ABSTRACT

Recently, pedestrian attributes like gender, age, clothing etc., have been used as soft biometric traits for recognizing people. Unlike existing methods that assume the independence of attributes during their prediction, we propose a multi-label convolutional neural network (MLCNN) to predict multiple attributes together in a unified framework. Firstly, a pedestrian image is roughly divided into multiple overlapping body parts, which are simultaneously integrated in the multi-label convolutional neural network. Secondly, these parts are filtered independently and aggregated in the cost layer. The cost function is a combination of multiple binary attribute classification cost functions. Experiments show that the proposed method significantly outperforms the SVM based method on the PETA database.

1. Introduction

Pedestrian attributes, such as *gender*, *dark hair*, and *skirt*, have been used as soft biometric traits in the surveillance field, and recently have attracted a lot of attention. For example, pedestrian attributes can be used as useful clues for person retrieval [2,3], person recognition [4,5] (also known as subject identification [6] and human identification [7–9]), face verification [10] and person re-identification [11]. In many real-world surveillance scenarios, cameras are usually installed at a far distance to cover wide areas, therefore pedestrians are captured with low resolutions. As a result, high-quality face images are hardly attainable. However, in such scenarios pedestrian attributes still have a high application potential, because pedestrian attributes have been shown to provide several advantages beyond traditional biometrics, such as invariance to illumination and contrast [6].

There are three main challenges in pedestrian attribute classification. First, there are large intra-class variations, due to diverse clothing appearances, various background conditions and different camera views. As shown in Fig. 1, the *backpack* annotated samples in the PETA [1] database captured with different cameras have drastic appearance variations. Second, pedestrian attributes have complex localizing characteristics, which means that some attributes can only be recognized in some certain or uncertain local body areas. For example, *long hair* is most relevant with the head and shoulders areas; *messenger bag* (see Fig. 1) may appear in either left of right side of the image, with uncertain heights. As a result, the pedestrian attribute feature extraction is very difficult. Third, pedestrian attribute classification is a multi-label classification problem instead of a multi-class classification problem, because pedestrian attributes are not completely mutually exclusive. Therefore, most of the existing multi-class classification algorithms are not applicable, and the multi-label classification problem has its own challenge.

The most popular approach for attribute prediction is the one using hand-crafted features and SVM based independent attribute classifier [1,6,11-13], which cannot solve the above-mentioned challenges successfully because hand-crafted features have limited representation ability for large intra-class variations, and independent SVM classifiers cannot investigate interactions between different attributes.

In this paper, we present a comprehensive study on pedestrian attribute classification. We propose a multi-label convolutional neural network (MLCNN) to solve the multi-attribute classification problem. The multi-label convolutional neural network is trained from raw pixels rather than hand-crafted features and is able to simultaneously recognize multiple attributes, which achieves higher accuracies than the SVM-based attribute classifiers proposed in [11–13]. The paper is built upon our preliminary work [14], which is improved from three aspects. The first one is that we train a deeper MLCNN and evaluate its performance on the PETA [1] database, which is the largest attribute pedestrian database to the best of our

☆ This paper has been recommended for acceptance by Michele Nappi.
* Corresponding author.
E-mail addresses: jqzhu@hqu.edu.cn (J. Zhu), scliao@nlpr.ia.ac.cn (S. Liao), zlei@nlpr.ia.ac.cn (Z. Lei), szli@nlpr.ia.ac.cn (S. Li).

http://dx.doi.org/10.1016/j.imavis.2016.07.004
0262-8856/

Fig. 1. Annotated sample images from the PETA [1] databases.

knowledge. The PETA database includes 19,000 pedestrian images, each of which is annotated with 65 attributes. The second one is that we do not connect each attribute with the corresponding pre-defined body parts as priors to help the network learning, since the PETA database holds a larger database size to drive the MLCNN to automatically learn discriminative features for attributes. The third one is that we propose a comprehensive evaluation protocol for pedestrian attribute classification, which not only reports classification accuracies, but also reports recall rates and areas under ROC curves (AUC).

2. Related work

2.1. Attribute pedestrian database

There are several public attribute pedestrian databases for surveillance applications, such as VIPeR [15], PRID [16], GRID [17], APiS [18], and PETA [1]. From the perspective of the number of attributes, VIPER is firstly annotated with 15 attributes by Layne et al. [11]. They annotated VIPER, PRID and GRID with 21 attributes in their further work [12]. The APiS database is annotated with 15 attributes by Zhu et al. [18]. PETA is the newest database, including 65 attribute annotations. From the perspective of the number of images, both VIPER, PRID and GRID are small databases and each one is less than 1500 images. The APiS and PETA databases include 3661 and 19,000 images, respectively. We can find that more and more databases are released, and the numbers of image and attribute annotation are increased. This situation illustrates that the study on pedestrian attribute classification is receiving more and more interests and attentions.

2.2. Pedestrian attribute classification

The most popular pedestrian attribute classification method is training each attribute classifier independently on hand-crafted features. In [6,11-13], each attribute classifier is trained by using a support vector machine (SVM). In [18], the gentle AdaBoost [19] algorithm is applied to train attribute classifier independently. These straightforward methods can train independently classifiers easily, if the number of attributes is small. However, when the number of attributes is huge, for example there are hundreds of attributes, the one by one training progress is too tedious for human. Moreover, these methods still have a room for improvement, because these methods ignore the interaction between different attributes.

There are some methods learning interaction models between different attributes to improve the performance of the pedestrian attribute classification. Chen et al. [20] explored the mutual dependencies between attributes by applying a conditional random field (CRF) with the SVM margins from the independently trained attribute classifiers. Deng et al. [1] exploited the context of neighboring images by an undirect graph based Markov random field (MRF), where each node represents a random variable and each edge represents the relation between the two connected nodes. Bourdev et al. [21] used the SVM algorithm to explore interactions between different attributes. Specially, they used the SVM algorithm learning on the prediction scores of all independently trained attribute classifiers to capture interactions between different attributes. In other words, the final decision score of an attribute is obtained by linearly combining all decision scores that come from independently trained attribute classifiers and the linear coefficients are learned by a SVM. However, since an attribute is most relevant to itself, the final decision score of an attribute in this interaction model will heavily rely on the decision score of its own attribute classifier, resulting in the role of other attributes is ignorable. In order to solve this disadvantage, Zhu et al. [22] improved the pedestrian attribute classification by weighted interactions from other attributes. In this method, the prediction of one attribute is achieved by a weighted combination of the independent decision score and the interaction score from other attributes. It is able to keep the balance of the independent decision score and interaction of other attributes to yield more robust classification results.

2.3. Convolutional neural network

The above-mentioned methods train attribute classiers on hand-crafted features. However, hand-crafted features have limited representation ability for large intra-class variations. Therefore, using machine learning based features is a potential improvement method. Convolutional neural networks (CNNs) [23–27] are very popular feature learning algorithms, which have been used in many image-related applications and exhibited good performances. The most relevant work is the multi-label deep convolutional ranking net proposed by Gong et. al [28] to address the multi-label annotation problem. Gong et. al [28] adopted the architecture proposed in [25] as basic framework and redesigned a multi-label ranking cost layer for multi-label prediction tasks.

3. Pedestrian attribute classifier training

3.1. Body part division

Because of body movements, commonly used holistic feature representation methods suffer from pose misalignments. Besides that, some attributes have local characteristic. For example, *long hair* is most relevant to head and shoulder areas; *backpack* is most likely to appear in upper torso regions; *jeans* appears in lower body parts. Considering these factors, in [20,29], a body part detection method is first applied to locate body regions and the corresponding features are produced by fusing all low-level features extracted from the detected regions. However, the body part detection itself, is a challenging problem, due to the geometric variation such as articulation and viewpoint changes as well as the appearance variation of the body parts arisen from versatile clothing types.

Since pedestrians are upright walking mostly, we do not use a body part detector to locate body parts accurately, but roughly divide a pedestrian image into multiple body parts with a sliding window strategy. As shown in Fig. 2, each pedestrian image is scaled into 128×48 pixels firstly. Then, a sliding window strategy is applied

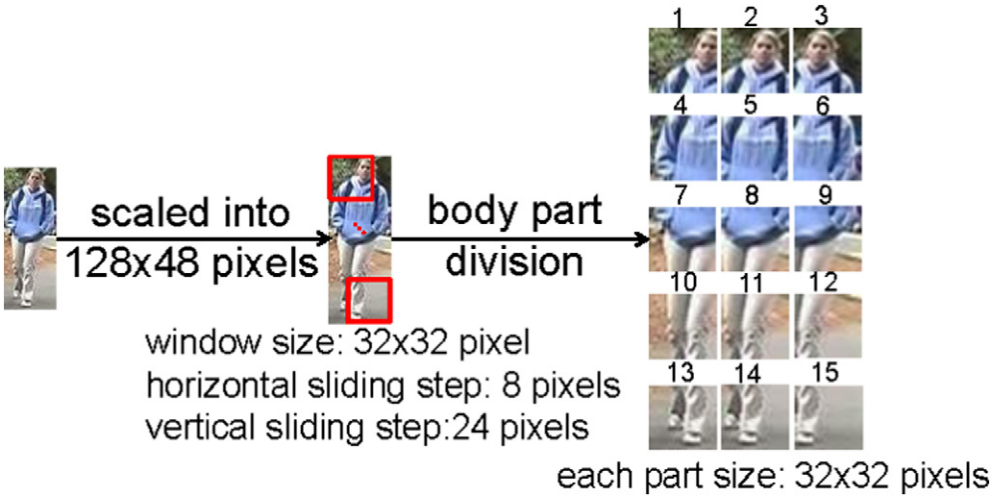

Fig. 2. One person is divided into 15 overlapping body parts.

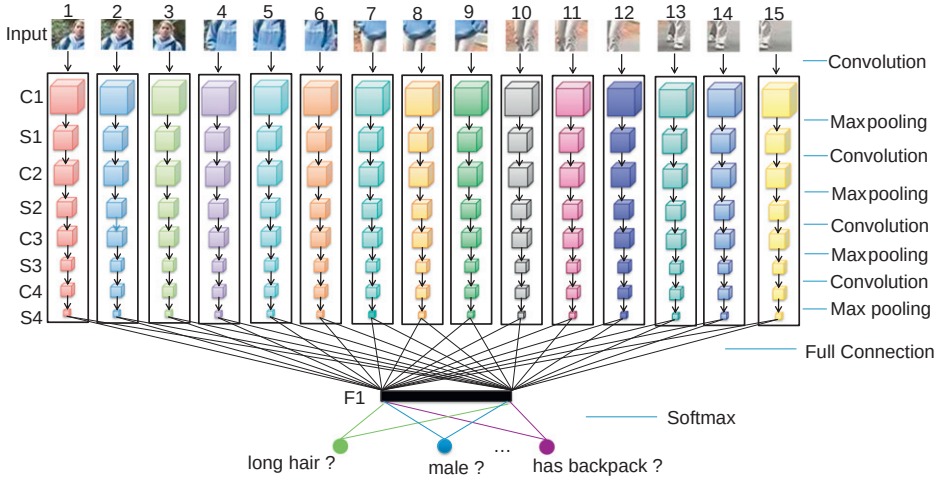

Fig. 3. The structure of the multi-label convolutional neural network (MLCNN).

to divide the scaled image into multiple equal sized body parts. In the sliding window strategy, the window size is 32 × 32 pixels, and the horizontal and vertical sliding steps are 8 pixels and 24 pixels, respectively. As a result, for a scaled pedestrian image, there are totally 15 overlapping body parts with 32 × 32 pixels, which are the inputs for the proposed MLCNN.

Table 1
The details of the proposed MLCNN.

Layer	Type	Output size	Neuron	Filter/Stride
C1	Convolution	32 × 32 × 16	ReLU	3 × 3/1
S1	Max pooling	16 × 16 × 16	–	3 × 3/2
C2	Convolution	16 × 16 × 32	ReLU	3 × 3/1
S2	Max pooling	8 × 8 × 32	–	3 × 3/2
C3	Convolution	8 × 8 × 48	ReLU	3 × 3/1
S3	Max pooling	4 × 4 × 48	–	3 × 3/2
C4	Convolution	4 × 4 × 64	ReLU	3 × 3/1
S4	Max pooling	2 × 2 × 64	–	3 × 3/2
F1	Full connection	256	ReLU	–

3.2. Multi-label convolutional neural network

After the body part division, multiple parts are integrated to a multi-label convolutional neural network (MLCNN) at the same time, as shown in Fig. 3. Each body part is filtered independently. The filter sizes of C1, C2, C3 and C4 layers are 3 × 3. The stride used in S1, S2, S3 and S4 are 2 pixels. All body parts are fully connected to the F1 layer to construct a feature representation. The ReLU neuron [25] is used as activation function for the convolution and full connection layers. The details of the proposed MLCNN are listed in Table 1.

3.3. Cost function and learning

Since attributes are not completely mutually exclusive, the prediction of multi-attribute is a multi-label classification problem essentially. The last layer of the proposed MLCNN structure is different from the CNN used for a single-label classification problem which usually only includes one cost function. In order to make our MLCNN to predict all attribute classifiers together, we sum all attribute classification cost functions together. Similar with [28,30], we use the

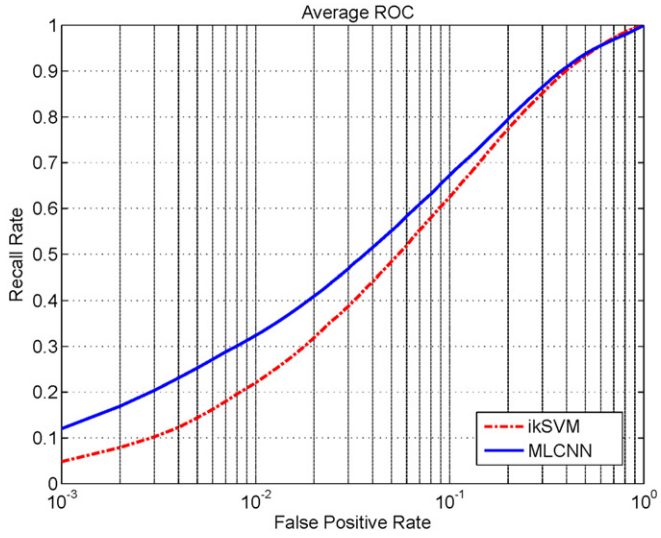

Fig. 4. The average ROC curve comparison between ikSVM [33] and our MLCNN on the PETA database.

softmax function [31] for the prediction of each attribute. The cost function of our multi-label convolutional neural network (MLCNN) is defined as follows:

$$F = \sum_{k=1}^{K} \lambda_k G_k \tag{1}$$

where G_k is the cost of the k-th attribute; K is the total number of the attributes; $\lambda_k \geq 0$ is a parameter used to control the contribution of the k-th attribute. In our experiments, we set $\lambda_k = \frac{1}{K}$ and define G_k as follows:

$$G_K = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M^k} 1\left\{y_n^k = m\right\} \cdot \log \frac{e^{\left(w_m^k\right)^T \cdot x_n^k}}{\sum_{m=1}^{M^k} e^{\left(w_m^k\right)^T \cdot x_n^k}}, \tag{2}$$

where $\left\{x_n^k, y_n^k\right\}$ represents a training sample and y_n^k is k-th attribute label of n-th sample x_n^k; N represents the number of training samples and M^k represents the class number of k-th attribute; $1\{\cdot\}$ is an indicator function. To avoid the bias due to imbalanced data, we further extend the Eq. (2) as follows:

$$G_k = -\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M^k} 1\left\{y_n^k = m\right\} \cdot \beta_m^k \cdot \log \frac{e^{\left(w_m^k\right)^T \cdot x_n^k}}{\sum_{m=1}^{M^k} e^{\left(w_m^k\right)^T \cdot x_n^k}}, \beta_m^k = \frac{\frac{1}{N_m^k}}{\sum_{l=1}^{M^k} \frac{1}{N_l^k}}, \tag{3}$$

where N_m^k is number of samples holding m-th class label of k-th attribute and it meets $\sum_{m=1}^{M^k} N_m^k = N^k$. Back propagation (BP) [23] is used to learn the parameters of the MLCNN and there are many public CNN learning tools, such as cudaconvnet [25] and Caffe [32].

4. Experiment and analysis

The newest challenging database PETA [1] is used to validate the superiority of our algorithm. The PETA database consists of 10 subsets, such as VIPER, PRID, GRID, and CAVIAR4REID, thus the PETA database is a complex database which includes different conditions, such as camera views, illuminations, resolutions and scenes. The PETA database includes 19,000 images and each image is annotated with 65 attributes, such as *gender*, *age*, *hair length*, and *clothing color*. Some attribute annotated samples are shown in Fig. 1. The hardware environment of following experiments is a notebook with i7-3720 QM CPU, NVIDIA Quadro K1000 M GPU and 16 GB memory.

4.1. Protocol

The evaluation protocol of [33] can be summarized as follows. 1) Each image in the PETA database is scaled into 128 × 48 pixels. 2) The PETA database is divided into non-overlapping training, validation and testing subsets, which includes 9500, 1900, and 7600 images, respectively. 3) Using the classification accuracy as the unique performance indicator of each attribute classification problem.

The classification accuracy is insufficient to evaluate the classification performance of an imbalanced attribute. To illustrate this shortcoming, an example is given. Suppose an imbalanced attribute which has 5% positive samples and 95% negative samples. For this attribute, even a naive classifier that determines all samples as negative ones will have a very high classification accuracy (95%). Therefore, besides of the classification accuracy, additional indicators are needed to fully evaluate the classification performance of an imbalanced attribute.

In order to overcome the shortcoming of the classification accuracy, we propose our evaluation protocol based on the aforementioned protocol as follows. Firstly, all multi-class attributes are transformed into binary class attributes. Secondly, each attribute's classification accuracy, recall rate when false positive rate (FPR) is set at 10% and area under the ROC curve (AUC) are reported. Finally, the average AUC of all attributes is also reported.

4.2. Setup

For the PETA database, a binary attribute is considered extreme imbalanced if the number of the corresponding positive sample is less than 500. We discard those extreme imbalanced attributes and obtain 45 binary attributes, as shown in Table 3. Both for training baseline ikSVM and our MLCNN models, the PETA database is augmented by the mirroring operation.

There are two baseline pedestrian attribute classification methods on the PETA database proposed in [33]. The first baseline method ikSVM is a SVM-based method. The features used for training ikSVM classifiers are the same with [12], which have 2784 dimensions, including 8 color channels such as RGB, HSV, and YCbCr, and 21 texture channels obtained by using the Gabor and Schmid filters on the luminance channel. The second baseline method MRFr2 [33] exploits the context of neighboring images by a Markov random field (MRF) to improve performance. The MRF is an undirect graph, where each node represents a random variable and each edge represents the relation between two connected nodes. The unary energy item is the probability predicted by ikSVM, while the pairwise energy item is similarly between neighboring images which learned by the random forest (RF) method. Table 2 lists the differences among the proposed MLCNN and the two baseline methods.

Since the experiments in [33] only report the classification accuracy of each attribute, we need to re-implement a baseline method

Table 2
Comparison of the proposed MLCNN and the two baseline methods [33].

Method	Feature type	Joint learning of all attributes	Using context modeling
MLCNN	CNN learning based	Yes	No
ikSVM [33]	Hand-crafted	No	No
MRFr2 [33]	Hand-crafted	No	Yes

Table 3
The performance comparison between ikSVM [33] and MLCNN on the PETA database. Bold font denotes the better case.

Attribute	Accuracy rate (%)		Recall rate (%) @ FPR=10%		AUC(%)	
	ikSVM	MLCNN	ikSVM	MLCNN	ikSVM	MLCNN
accessoryHat	92.04	**96.05**	81.37	**86.06**	91.27	**92.62**
accessoryMuffler	94.84	**97.17**	**90.68**	88.42	**95.09**	94.47
accessoryNothing	78.87	**86.11**	35.37	**52.57**	81.79	**86.09**
carryingBackpack	76.39	**84.30**	46.19	**58.40**	84.52	**85.19**
carryingMessengerBag	74.51	**79.58**	50.22	**58.30**	78.44	**82.01**
carryingNothing	75.84	**80.14**	49.36	**55.15**	81.60	**83.08**
carryingOther	76.18	**80.91**	38.57	**46.90**	74.11	**77.68**
carryingPlasticBags	86.86	**93.45**	**70.57**	67.30	**87.69**	86.01
footwearBlack	74.29	**75.97**	50.37	**57.24**	81.42	**84.07**
footwearBrown	82.38	**92.14**	63.19	**65.77**	84.67	**85.26**
footwearGrey	79.34	**87.07**	48.78	**50.80**	**80.93**	80.92
footwearLeatherShoes	81.89	**85.26**	66.58	**72.28**	87.33	**89.84**
footwearShoes	72.34	**75.78**	46.91	**52.80**	79.16	**81.63**
footwearSneakers	78.01	**81.78**	45.51	**52.04**	**83.32**	83.19
footwearWhite	78.99	**85.89**	52.34	**62.72**	83.84	**86.16**
hairBlack	84.76	**87.83**	75.54	**81.03**	91.88	**93.61**
hairBrown	84.24	**89.58**	72.24	**77.36**	89.77	**91.33**
hairGrey	92.18	**95.25**	71.05	**74.91**	87.82	**89.42**
hairLong	79.26	**88.12**	55.99	**76.49**	84.18	**90.55**
hairShort	77.64	**86.93**	52.48	**69.68**	82.90	**89.84**
lowerBodyBlack	**84.54**	83.86	**75.56**	71.21	**91.77**	90.84
lowerBodyBlue	85.64	**88.64**	72.42	**77.26**	90.15	**90.81**
lowerBodyCasual	85.47	**90.54**	53.66	**56.23**	85.60	**87.49**
lowerBodyFormal	84.63	**90.86**	65.42	**72.52**	85.99	**87.79**
lowerBodyGrey	78.66	**82.07**	**54.48**	53.43	**84.11**	82.77
lowerBodyJeans	78.58	**83.13**	57.22	**67.59**	84.97	**87.71**
lowerBodyTrousers	73.41	**76.26**	49.66	**56.19**	80.65	**84.16**
personalLarger60	96.08	**97.58**	89.09	**90.71**	**95.34**	94.94
personalLess30	79.34	**81.05**	61.27	**63.75**	86.73	**88.50**
personalLess45	76.09	**79.87**	51.14	**59.42**	82.11	**84.62**
personalLess60	79.71	**92.84**	64.69	**70.22**	84.94	**87.66**
personalMale	78.45	**84.34**	54.05	**74.80**	85.79	**91.74**
upperBodyBlack	85.80	**86.21**	**81.35**	80.11	**93.23**	93.06
upperBodyBlue	92.75	**94.53**	**80.61**	76.19	**92.95**	90.92
upperBodyBrown	89.41	**93.25**	**72.06**	68.60	**88.94**	87.58
upperBodyCasual	81.83	**89.25**	47.11	**62.14**	83.71	**87.17**
upperBodyFormal	87.11	**91.12**	62.42	**70.48**	85.22	**87.57**
upperBodyGrey	82.43	**84.39**	**60.38**	55.33	**85.61**	82.99
upperBodyJacket	88.75	**92.34**	**53.93**	53.37	**83.33**	80.98
upperBodyLongSleeve	84.80	**87.88**	**76.50**	74.29	**90.92**	89.97
upperBodyOther	79.67	**81.97**	70.09	**73.19**	87.05	**88.50**
upperBodyRed	95.61	**96.33**	**90.86**	86.77	**96.58**	94.69
upperBodyShortSleeve	83.05	**88.09**	68.20	**69.22**	**89.91**	89.21
upperBodyTshirt	84.13	**90.59**	**63.81**	63.51	**89.31**	88.73
upperBodyWhite	87.00	**88.84**	**76.18**	75.25	**92.29**	91.24
Average	82.75	**87.23**	62.57	**67.29**	86.42	**87.66**

for evaluation of the new performance measure. As shown in [33], the MRFr2 is only slightly better than the ikSVM algorithm but these is no open-source code of MRFr2 available, therefore, we re-implement the ikSVM method based on the feature extraction and ikSVM training codes released by the authors of [12] and [34], respectively.

4.3. Attribute classification

Following the proposed evaluation protocol, we report the performances of ikSVM [33] and our MLCNN methods, as shown in Table 3 and Fig. 4. Each attribute classifier is with the default threshold, that is 0 for the ikSVM and 0.5 for the MLCNN, respectively. For classification accuracies, it can be found that our MLCNN offers higher classification accuracies for 44 of 45 attributes. The average accuracy of MLCNN achieves 87.23% and it is 4.48% higher than that of the ikSVM method [33]. For the recall rates at FPR=0.1, our MLCNN model outperforms the ikSVM method for 34 of 45 attributes and obtains 4.72% higher a average recall rate. For AUC performances, our MLCNN method obtains larger AUCs for 31 of 45 attributes and also achieves a larger average AUC. From Fig. 4, we can clearly find that the average ROC of our MLCNN is better than that of the ikSVM method proposed in [33].

These results illustrate that our MLCNN method achieves better classification performances for most attributes, but it is beaten by the ikSVM method for a few remaining attributes. The reason is that the ikSVM and our MLCNN classifiers are trained under different principles. The ikSVM classifiers are independently trained for different attributes, which means that each ikSVM classifier is independently optimized for the classification of the corresponding attribute. On the contrary, the proposed MLCNN method learns all attribute classifiers together, which is optimized for the overall classification performance of all attributes. As a result, the proposed MLCNN method is able to obtain a better overall classification performance of all attributes, but loses superiorities for a few attributes.

5. Conclusion

In this paper, a multi-label convolutional neural network (MLCNN) for the pedestrian attribute classification problem is proposed. The multi-label convolutional neural network is trained from

raw pixels rather than hand-crafted features and is able to simultaneously predict multiple attributes. Experimental results on the PETA database have well demonstrated the superiority of the MLCNN pedestrian attribute classification method.

The future work will be done in two directions. The first one is to develop an effective method to adaptively adjust the weight of each attribute in the multi-label cost function. This is to ensure that each attribute receives enough attention, so as to avoid the situation that the costs of some attributes are less addressed during the MLCNN training process. The second one is to develop a novel method to explore the interaction among different attributes to improve the classification performance.

Acknowledgments

This work was supported in part by the Scientific Research Funds of Huaqiao University under the Grant 16BS108, in part by the National Natural Science Foundation of China under the Grants 61602191, 61375037, 61473291, 61572501, 61502491 and 61572536, in part by the Chinese Academy of Sciences Project under the Grant KGZD-EW-102-2.

References

[1] Y. Deng, P. Luo, C.C. Loy, X. Tang, Pedestrian attribute recognition at far distance, International Conference on Multimedia, ACM. 2014, pp. 789–792.
[2] E.S. Jaha, M.S. Nixon, Analysing soft clothing biometrics for retrieval, Biometric Authentication Springer. 2014, pp. 234–245.
[3] A. Dantcheva, A. Singh, P. Elia, J. Dugelay, Search pruning in video surveillance systems: efficiency-reliability tradeoff, International Conference on Computer Vision Workshops, IEEE. 2011, pp. 1356–1363.
[4] A.K. Jain, S.C. Dass, K. Nandakumar, Soft biometric traits for personal recognition systems, Biometric Authentication Springer. 2004, pp. 731–738.
[5] A. Dantcheva, J.-L. Dugelay, P. Elia, Person recognition using a bag of facial soft biometrics (BoFSB), International Workshop on Multimedia Signal Processing, IEEE. 2010, pp. 511–516.
[6] E.S. Jaha, M.S. Nixon, Soft biometrics for subject identification using clothing attributes, International Joint Conference on Biometrics, IEEE. 2014, pp. 1–6.
[7] D.A. Reid, M.S. Nixon, S.V. Stevenage, Soft biometrics; human identification using comparative descriptions, Transactions on pattern analysis and machine intelligence 36 (6) (2014) 1216–1228.
[8] E. Martinson, W. Lawson, J.G. Trafton, Identifying people with soft-biometrics at fleet week, International Conference on Human-Robot Interaction, IEEE. 2013, pp. 49–56.
[9] A. Dantcheva, J.-L. Dugelay, P. Elia, Soft biometrics systems: reliability and asymptotic bounds, International Conference on Biometrics: Theory Applications and Systems, 2010. pp. 1–6.
[10] N. Kumar, A.C. Berg, P.N. Belhumeur, S.K. Nayar, Attribute and simile classifiers for face verification, Conference on International Conference on Computer Vision, IEEE. 2009, pp. 365–372.
[11] R. Layne, T.M. Hospedales, S. Gong, Person re-identification by attributes, British Machine Vision Conference, 2012. pp. 8.
[12] R. Layne, T.M. Hospedales, S. Gong, Attributes-based re-identification, Person Re-Identification Springer. 2014, pp. 93–117.
[13] L. An, X. Chen, M. Kafai, S. Yang, B. Bhanu, Improving person re-identification by soft biometrics based reranking, International Conference on Distributed Smart Cameras, IEEE. 2013, pp. 1–6.
[14] J. Zhu, S. Liao, D. Yi, Z. Lei, S.Z. Li, Multi-label CNN based pedestrian attribute learning for soft biometrics, International Conference on Biometrics, IEEE. 2015, pp. 535–540.
[15] D. Gray, S. Brennan, H. Tao, Evaluating appearance models for recognition, reacquisition, and tracking, IEEE International Workshop on Performance Evaluation for Tracking and Surveillance.
[16] M. Hirzer, C. Beleznai, P.M. Roth, H. Bischof, Person re-identification by descriptive and discriminative classification, Image Analysis Springer. 2011, pp. 91–102.
[17] C. Liu, S. Gong, C.C. Loy, X. Lin, Person re-identification: what features are important? European Conference on Computer Vision Workshops, Springer. 2012, pp. 391–401.
[18] J. Zhu, S. Liao, Z. Lei, D. Yi, S.Z. Li, Pedestrian attribute classification in surveillance: database and evaluation, International Conference on Computer Vision Workshops, IEEE. 2013, pp. 331–338.
[19] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), The Annals of Statistics 28 (2) (2000) 337–407.
[20] H. Chen, A. Gallagher, B. Girod, Describing clothing by semantic attributes, European Conference on Computer Vision Springer. 2012, pp. 609–623.
[21] L. Bourdev, S. Maji, J. Malik, Describing people: a poselet-based approach to attribute classification, International Conference on Computer Vision, IEEE. 2011, pp. 1543–1550.
[22] J. Zhu, S. Liao, Z. Lei, S.Z. Li, Improve pedestrian attribute classification by weighted interactions from other attributes, Asian Conference on Computer Vision Workshops, Springer. 2014, pp. 545–557.
[23] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.
[24] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, International Conference on Machine Learning, ACM. 2009, pp. 609–616.
[25] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, 2012. pp. 1097–1105.
[26] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, A. Senior, P. Tucker, K. Yang, Q.V. Le, Large scale distributed deep networks, Advances in Neural Information Processing Systems, 2012. pp. 1223–1231.
[27] G.E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R.R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.
[28] Y. Gong, Y. Jia, T. Leung, A. Toshev, S. Ioffe, Deep convolutional ranking for multilabel image annotation, arXiv preprint arXiv:1312.4894.
[29] A. Li, L. Liu, K. Wang, S. Liu, S. Yan, Clothing attributes assisted person reidentification, Transactions on Circuits and Systems for Video Technology 25 (5) (2015) 869–878.
[30] M. Guillaumin, T. Mensink, J. Verbeek, C. Schmid, Tagprop: discriminative metric learning in nearest neighbor models for image auto-annotation, International Conference on Computer Vision, IEEE. 2009, pp. 309–316.
[31] C.M. Bishop, Pattern Recognition and Machine Learning, 1. springer. 2006.
[32] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: convolutional architecture for fast feature embedding, arXiv preprint arXiv:1408.5093.
[33] Y. Deng, P. Luo, C.C. Loy, X. Tang, Learning to recognize pedestrian attribute, arXiv preprint arXiv:1501.00901.
[34] S. Maji, A.C. Berg, J. Malik, Classification using intersection kernel support vector machines is efficient, Conference on Computer Vision and Pattern Recognition, IEEE. 2008, pp. 1–8.