ISSN 1004-9037,CODEN SCYCE4
Journal of Data Acquisition and Processing Vol. 33,No. 4,Jul. 2018,pp. 751－757
DOI:10.16337/j.1004-9037.2018.04.019
© 2018 by Journal of Data Acquisition and Processing
http://sjcj.nuaa.edu.cn
E-mail:sjcj@nuaa.edu.cn
Tel/Fax:＋86-025-84892742

基于深度学习与属性学习相结合的行人再识别

许方洁　成科扬　张建明

(江苏大学计算机科学与通信工程学院,镇江,212013)

摘　要: 现实情况中缺少大量有标签数据,导致有监督的行人再识别模型训练受到影响。此外,低层特征的缺乏语义特性限制了行人再识别在行人检索、罪犯追踪等中的应用。本文提出了一种基于深度学习与属性学习相结合的行人再识别方法,利用深度学习的无监督模型提取行人图像的本质特征,并引入"属性"概念增强特征的语义表达能力。首先采用卷积自动编码器进行无监督的特征提取,提取的特征然后交由多个属性分类器进行属性分类,并结合统计获得的属性类别映射关系表计算最终类别判定,最后在 VIPeR 和 i-LIDS 标准数据集上进行了测试,并与基于优化属性的行人再识别方法(Optimized attribute based re-identification,OAR)、显著性检测对应法(Salience detection correspondence,SDC)等进行了比较,结果表明本方法能够赋予行人再识别较好的语义性能,并在一定程度上提高了识别的准确率,同时获得了较好的零训练样本识别效果。

关键词: 行人再识别;深度学习;卷积自动编码器;属性学习

中图分类号: TP391.41　　**文献标志码:** A

Pedestrian Re-identification Based on Deep Learning and Attributes Learning

Xu Fangjie, Cheng Keyang, Zhang Jianming

(School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang,212013, China)

Abstract: The lack of much labeled data in the real world affects the training of supervised model for pedestrian re-identification. Besides, applications of pedestrian re-identification in pedestrian retrieving and criminal tracking are limited because of the lack of semantic representation. This paper presents a pedestrian re-identification method based on the combination of deep learning and attributes learning, which extracts essential features with unsupervised deep learning model and enhances the semantic representation of features with 'attributes'. Firstly, a convolutional auto-encoder (CAE) is used to extract features of unlabeled pedestrian images, and the extracted features are then input into several attribute classifiers to judge whether the pedestrian owns the attributes. Lastly, with a table of 'attributes-classes mapping relations', we can get the final dassification result. Tests of the proposed algorithm and comparisons with other algorithms on the VIPeR and i-LIDS datasets are shown, and results prove that our algorithm indeed strengthens the semantic representation and improves the accuracy of pedestrian re-identification, achieving good 'zero-shot' re-identification performance as well.

Key words: pedestrian re-identification; deep learning; convolutional auto-encoder (CAE); attribute learning

---

基金项目: 江苏省普通高校研究生科研创新计划(KYLX15_1080)资助项目;国家自然科学基金(61602215)资助项目;江苏省自然科学基金(BK20150527)资助项目;镇江市科技计划(SH2014017)资助项目;江苏大学高级人才科研启动基金(15JDG180)资助项目。

收稿日期: 2016-09-18;**修订日期:** 2016-10-28

引　言

随着人工智能的发展和高清监控摄像的普及，人们的公共安全意识逐步提高，行人视频监控已经渗透在我们生活的各个方面。行人再识别作为行人视频监控中最为重要的一项技术，广泛应用于类似公共场所监控、犯罪预防和案件侦破等安全领域。行人再识别，即在某摄像头中出现的人物在另一摄像头中出现或在同一摄像头中不同时间出现时，计算机能够自动判断是同一人物的技术，主要由特征提取与分类器设计两大模块构成。在特征表示方面，颜色、纹理、边缘和形状等传统特征被广泛地应用在行人特征的表示上，或将多种具有互补性质的特征相结合[1]，在一定程度上来克服因视角、光线等条件变化造成的影响。然而，这些特征并非都具有同等地位，很有必要对不同特征进行加权，文献[2]提出无监督显著性学习方法，其通过学习每个特征块的显著性来对两幅图的相似性进行加权，有效提高了算法的再识别率；文献[3]在 Zhao 等工作的基础上，针对行人再识别中目标类间显著性外观特征的不稳定问题，融合图像的内在显著性特征提出了一种新的行人再识别算法。特征分布的空间信息也是一种重要的线索，将行人图像分割成不同片段或区域来进行特征提取[4,5]，分割方法有横向条纹分割、三角图形分割、同心环分割及局部块分割等。基于人体结构对称这一生物特性，文献[6]将行人分为头部、躯干和腿部，分别以躯干和腿部的垂直对称轴为中心建立基于国际照明学会推荐颜色空间模型(Commission international eclairage LAB，CIELAB)的局部加权空间直方图，并结合贝叶斯框架下基于局部统计特征的显著区域检测方法描述行人外观特征。

以上人为设定的特征受到场景影响较大，通常需要专业人员针对不同场景进行深入研究来确定[7]。此外，人体作为一种非刚性对象，采用人工设定的固定特征来表示行人并非是最好的选择。从机器学习的角度来看，由数据驱动提取的特征更能够体现行人对象的本质以及行人之间的区分性。深度学习是近年来机器学习领域较为热门的话题，它是一种特征学习方法，把原始数据通过一些简单的非线性模型转变为更高层次的、更加抽象的表达[8]。相比传统神经网络，其优势在于通过无监督的逐层模型来感知图像的高层特征，从而获得更能表示图像的本质特征。此外，行人视频监控领域中存在的大量无标签数据，对于传统有监督的学习方法来说是影响识别效率的重大问题，而深度学习中的无监督算法恰恰适应了这种情况，为行人再识别带来了希望。

将提取到的图像特征输入分类器可获得从低层特征到类别的直接映射，然而在行人再识别中，获得每个行人的大量训练样本是不可能实现的，因此，这种传统低层特征直接到类别的映射方法在行人再识别任务中并不可取。此外，基于数字的低层特征并不能被人所理解，其缺乏语义的本质限制了行人再识别的应用，例如，在对罪犯的所搜中，目击者往往只能提供罪犯的发型、服装和携带物等外貌特征。而由于监控视频清晰度有限或者仅仅捕捉到行人的侧脸或后背等原因，发展成熟的人脸识别技术常常没有办法很好地应用在行人再识别上。相比于人脸上的眼睛、鼻子等细小的五官特征，在监控视频中更容易获得发型、服装这类明显的外貌特征。这些外貌特征称之为行人的属性，即可用于区分行人的具有语义表达能力的中层特征。属性学习就是在低层特征与类别之间加入一层具有语义性质的属性，使得行人识别算法更符合实际应用需求。本文结合了深度学习与属性学习的优势，将两者结合并应用于行人再识别该特定任务中，从特征的提取与表示两方面提高了行人再识别算法的准确性与实用性。

1　基于深度学习与属性学习相结合的行人再识别方法

1.1　深度学习与属性学习的结合

本文提出的基于深度学习与属性学习相结合的行人再识别方法主要由两个模块组成：特征提取模块与属性分类模块。特征的提取选用深度学习中的无监督学习模型——卷积自动编码器(Convolution-

al auto-encoder, CAE)来对卷积神经网络(Convolutional neural network, CNN)模型进行训练,解决了无标签样本的训练问题。由属性分类器对 CAE 提取的特征进行属性分类,并结合"属性-类别映射关系"获得最终的识别结果。本文的行人再识别方法整体框架参考图 1。

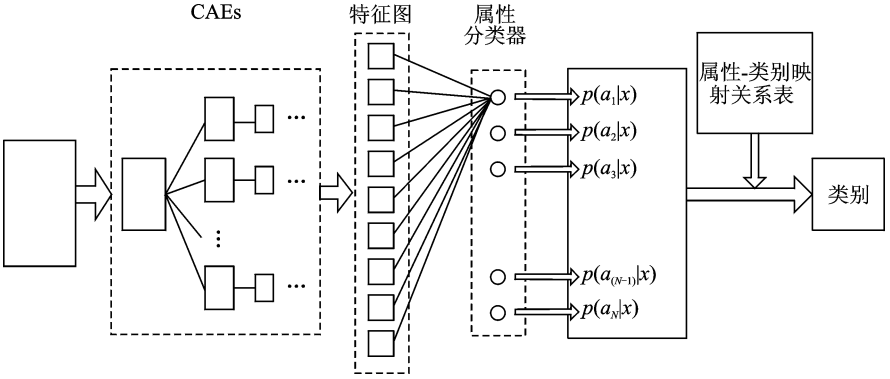

图 1 基于深度学习与属性学习相结合的行人再识别方法框架

Fig. 1 Framework of deep learning and attributes learning based pedestrian re-identification method

输入图片经过多层 CAE 的特征提取获得多张特征图,属性分类器集合中包含多个属性相关的支持向量机(Support vector machine, SVM)分类器,将 CEA 提取的多张特征图拉成一条向量并进行稀疏化,因此,每个输入样本对应一个稀疏向量。将该稀疏向量输入所有分类器,获得对应属性的概率 $p(a_k \mid x)$ 分布,即对于每个属性,样本具有该属性的概率。"属性-类别映射关系表"由统计而来,其中每个元素代表具有属性 a_k 的情况下属于类别 y_j 的概率。

1.2 卷积自动编码器

普通自动编码器与堆叠自动编码器不能处理二维图像的池化与白化问题,并且大量冗余参数被强迫参与计算使得运算效率低下,而 CAE 是一种专门用于处理图像数据的神经网络[9],具有无监督学习的能力,能够为卷积神经网络提供模型训练。本文针对所选数据集为行人再识别设计了一个多层 CNN 网络,并通过 CAE 对每一层网络进行训练,过程如图 2 所示。每一层的输入图像 X 进行卷积操作得到一系列特征图 h,经过 Max-Pooling 后通过反卷积操作重构原输入图像得到 Y,X 与 Y 的误差作为该隐含层的参数调整依据,权值和偏置的更新通过传统反向传播算法(Backpropagation algorithm, BP)获得。参考文献[10],根据数据集图像构建 3 层隐层网络:第 1 层卷积层 C_1 设定 100 张特征图,为集合 h_1,卷积核 K_1 大小为 5×5,次采样层采用 2×2 的 Max-Pooling 窗口;第 2 层卷积层 C_2 设定 150 张特征图,集合为 h_2,卷积核 K_2 大小为 5×5,次采样层 Max-Pooling 窗口大小为 2×2;第 3 层卷积层 C_3 特征图有 200 张,集合为 h_3,卷积核 K_3 大小针对图像尺寸设为 3×3,次采样层 Max-Pooling 窗口大小为 2×2。卷积操作步长为 1,Max-Pooling 操作步长均为 2。经过 Max-Pooling 后的图像保持原来尺寸,仅把 Max-Pooling 窗口扫过区域中非最大值的像素设为 0,而 Pooling 操作后,图像中的 0 像素均被删除,原图像尺寸减半。

1.3 属性学习

1.3.1 行人属性

行人属性是关于行人外貌的语义性描述,人体不同的部位具有不同的属性,例如,人体头部相关属性有"秃顶"、"带帽"、"长发"以及"短发"等;衣服款式相关属性有"长袖""短袖""连衣裙"和"短裙"等;鞋子相关属性有"单色鞋""多色鞋""平底鞋"和"高跟鞋"等;携带物相关属性有"双肩包""单肩包""手提

包”和“无携带”等。在不同环境和场合可以对行人属性进行选择和细分，使之有利于行人的再识别。例如，在机场、火车站等流通区域，“行李箱”“背包”等属性就成为普遍属性，而这些属性在商场可能并不适用。本文参考行人再识别领域专家的研究，并基于 VIPeR 数据库行人特点，选取数据集中存在并具有区分性的属性用于行人再识别。具体选用了“带帽”“短发”“长发”“长袖”“短袖”“穿外套”“长裤”“短裤”“有花纹”“有 LOGO”“双肩包”“单肩包”“手提包”“手拿物品”“凉鞋”“单鞋”和“靴子”这 17 个属性。对所有训练样本进行关于这些属性的标记，并为每一个属性训练一个 SVM 分类器。图片的属性标签以 1 和－1 表示，1 代表图片中行人具有该属性，为正样本，而－1 则代表不具有该属性，为负样本。图 3 展示了部分具有属性的样本。

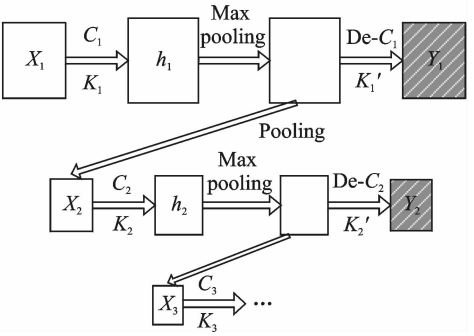

图 2 卷积自动编码器编码器训练过程

Fig. 2 Training process of CAEs

(a) 拥有“双肩包”属性的行人样本

(a) Samples with ‘backpack’ attribute

(b) 拥有“长发”属性的行人样本

(b) Samples with ‘long hair’ attribute

图 3 行人属性示例

Fig. 3 Examples of pedestrian’s attributes

1.3.2 属性学习

对于有监督的学习方法，每个类别需要大量样本来对模型进行训练，而在实际情况中，搜集每个行人的大量图像并不可取，而属性学习则能够有效克服类别样本稀缺问题。由于在属性学习阶段，需要知道的是某个行人是否具有某个属性，而不是该行人是谁，而属性具有共享性，即某个属性被许多人所共有，那么针对属性的学习就可以利用这些具有该属性的不同行人样本进行训练，而具有该属性的训练样本就可以很充足，足以训练出效果良好的分类器。

行人属性设计完成后，为每一个属性设计一个 SVM 分类器，将 CAE 最后获得的特征作为属性分类器的输入。训练分类器时，具有该属性的样本作为正样本，不具有该属性的样本作为负样本，将正负样本输入该属性分类器以进行训练，以获得样本具有该属性的概率，即属性的后验概率 $p(a_k \mid x)$，每个属性分类器都作类似训练。属性的先验概率由统计获得，构成一张“属性-差别”映射概率表。表 1 中展示了 10 个类别对应 10 个属性的概率，每一元素代表具有属性 a_k 的样本属于类别 y_j 的概率，即 $p(y_j \mid a_k)$。根据条件概率公式计算类别的后验概率，即

$$p(y_j \mid x_t) = \sum_{k=1}^{N} p(y_j \mid a_k) p(a_k \mid x_t) \tag{1}$$

取类别后验概率最大的类作为对样本的判别类，即

$$\hat{y} = \arg\max_{y_j} p(y_j \mid x_t) \tag{2}$$

表 1 属性-类别映射概率表示例

Tab. 1 Example of the 'attributes-classes mapping propability' table

类别	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
y_1	0.00	0.11	0.00	0.00	0.14	0.00	0.00	0.33	0.00	0.13
y_2	0.00	0.11	0.00	0.00	0.14	0.00	0.06	0.00	0.00	0.13
y_3	0.00	0.00	0.14	0.08	0.00	0.13	0.06	0.00	0.00	0.00
y_4	0.00	0.11	0.00	0.00	0.14	0.00	0.00	0.33	0.33	0.33
y_5	0.00	0.11	0.00	0.08	0.00	0.13	0.06	0.00	0.00	0.00
y_6	0.00	0.00	0.14	0.08	0.00	0.13	0.06	0.00	0.00	0.00
y_7	0.17	0.00	0.00	0.08	0.00	0.13	0.06	0.00	0.00	0.00
y_8	0.00	0.11	0.00	0.00	0.14	0.00	0.06	0.00	0.00	0.00
y_9	0.00	0.00	0.14	0.00	0.14	0.00	0.06	0.00	0.00	0.00
y_{10}	0.17	0.00	0.00	0.00	0.14	0.13	0.06	0.00	0.00	0.00

2 实验结果与分析

2.1 数据库

本文采用行人再识别中常用的 VIPeR 公共行人数据库[11]进行训练和测试。VIPeR 数据库包含 632 个行人的 1 264 张图像,由两个不同视场的摄像头拍摄,每个行人分别具有两张不同角度、光线和姿态的图片。对于有监督的学习方法,每个行人 2 张图片无法训练有效的识别模型,而属性学习成功避免了这个问题。根据本文设定的 17 个属性对 1 264 张样本图片进行属性标注,图 4 给出了其中 1 张行人图像所具有的属性,这些属性对应的标签即为 1,其余属性标签即为−1。632 个行人对应了 632 个类别,实验中,图片样本类别以 1～632 的数字来表示。本文采用文献[12]提出的交叉验证方法,将 VIPeR 数据库随机分为 10 等份,1 份作为测试集,确保其中不存在属性完全相同的两个样本,其余 9 份用于训练,重复 10 次,以 10 次实验的平均值作为实验结果。

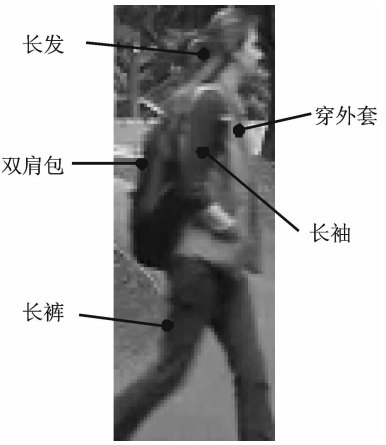

图 4 行人图片对应属性

Fig. 4 Corresponding attributes in one pedestrian sample

2.2 实验结果与分析

经过 10 次实验,对结果求平均,以上 17 个属性分类器的测试准确率如图 5 所示,可以看出数据集对于某些属性,如"短发""长发""长袖"和"短袖"等属性的分类准确率相比其他属性稍低。这是由于 VIPeR 数据集中采集的行人图片背景比较复杂,尤其上半身区域对应背景中常有树木、建筑等干扰图像,若对图片进行适当的背景分离,属性分类器的识别效率应该会更高。

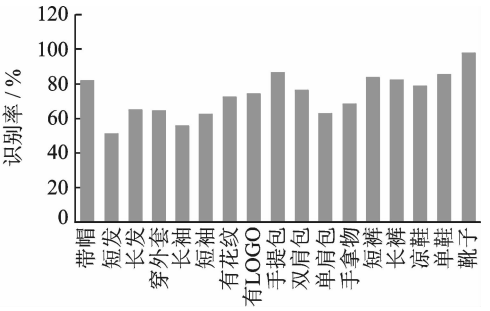

图 5 属性分类器识别率

Fig. 5 Accuracies of attribute classifiers

本文采用的 17 个属性与文献[13]所采用的属性有部分相同或相似,表 2 列举了 6 个相同及相似属性分类器的分类准确率以及该 6 个属性分类器的平均准确率作为比较。文献[13]采用颜色和纹理这两个传统的低层特征作为属性分

类器的输入，颜色特征包括 RGB，HSV 和 YCbCr 共 8 个通道，再由亮度通道导出 21 个纹理特征，随后将特征交给 SVM 分类器进行分类识别。由表 2 可知，对于相同属性或相似属性，由本文 CAE 提取的特征所训练的属性识别准确率总体上要高于文献[13]中基于优化属性的行人再识别方法(Optimized attribute based re-identification，OAR)方法达到的准确率，如在属性“有 logo”上识别率高出 10%以上，而在属性“手提包”上高出 30%以上，在属性“双肩包”、“穿短裤”和“穿长裤”上识别率也有较大的提高。并且，属性的平均识别率也有所提高，由此体现了深度学习模型 CAE 在特征提取上的优势。

如表 3 所示，本文在 VIPeR 数据集上进行行人再识别实验的准确率达到 27.5%，与文献[13]所达到的最高第一匹配率 21.4%相比高出约 6%。与文献[14]采用的设定属性与自挖掘属性结合的方法相比，识别率高出近 10%。与文献[6]采用显著性检测对应 k 最近邻法(Salience detection correespondence k-nearest neighbor，SDC-knn)和显著检测对应单类支持向量机法(Salience detection correspondence one-class support vector machine, SDC-OCSVM)对颜色特征与尺度不变特征变换(Scale-invariant feature transform，SIFT)特征进行分类的方法相比，本文识别率也占有优势。文献[15]采用改进的深度学习方法进行特征提取，实现了 28.2%的识别率，虽然高于本文方法，但其方法提取的特征缺乏语义表达能力，因此在实际应用中缺乏实用价值。

表 2 属性分类器准确率比较

Tab. 2 Comparison of accuracies of attribute classifiers %

属性	OAR 方法	本文方法
外套	69.7	64.3
有 LOGO	60.8	74.6
手提包	54.5	86.5
双肩包	68.6	76.2
短裤	76.1	84.1
长裤	78.0	81.7
平均	68.0	78.0

表 3 在 VIPeR 数据库上的行人再识别识别率

Tab. 3 Recognition rates on VIPeR data set %

方法	识别率
本文方法	27.5
OAR 方法[13]	21.4
Umeda 等的方法[14]	18.0
SDC_knn 方法[6]	26.3
SDC_ocsvm 方法[6]	26.7
Ahmed 等的方法[15]	28.2

由于属性具有共享性，本文所提出的行人再识别方法同样可以对于缺乏训练样本的行人图像进行识别区分。将本文训练好的深度模型和属性分类器直接用于 i-LIDs[16] 行人数据集，得到 22.1%的识别率。由表 4 可知，文献[17]提出的基于属性解释的再识别方法(Attribute interpreted-identification，AIR)以及加权 AIR 方法在 i-LIDS 上的零训练样本识别率分别为 11.5%和 16.5%，均低于本文所达到的零训练样本识别率。

表 4 在 i-LIDS 数据集上的零样本识别率

Tab. 4 Zero-shot recognition rates on i-LIDS data set %

方法	识别率
本文方法	22.1
AIR 方法[17]	11.5
W. AIR 方法[17]	16.5

实验证明，本文提出的基于深度学习与属性学习相结合的行人再识别方法能够有效提高行人再识别的准确率，并增强再识别系统的语义表达能力，同时，对于零训练样本能够实现较好的识别能力，使得该行人再识别系统能够更有效、更实用地应用在各安全领域。

3 结束语

行人再识别具有巨大发展前景，但也存在着因视角、姿态、光照和遮挡等各种变化因素导致的挑战。本文提出了一种基于深度学习与属性学习相结合的行人再识别方法，首先，为行人图像构建一个多层卷积神经网络用以提取行人特征，并通过卷积自动编码器进行网络的无监督训练；然后设计了 17 个行人外貌相关属性来对不同的行人进行区分，充分利用了深度学习提取特征的优势与属性富有语义的特点，

在一定程度上提高了存在各种变化因素情况下行人再识别方法的准确率与实用性。得益于属性所具有的共享特性,本文提出的方法能够对缺乏训练样本的行人图像进行较好的迁移识别能力。由于属性的设计对再识别的影响重大,如何实现数据驱动的属性生成方法是下一步研究的主要内容。

参考文献:

[1] Farenzena M, Bazzani L, Perina A, et al. Person re-identification by symmetry-driven accumulation of local features[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. San Francisco,USA: IEEE, 2010: 2360-2367.

[2] Freeman W T, Roth M. Orientation histograms for hand gesture recognition[C]// International Work-shop on Automatic Face and Gesture Recognition. Zurich,Swiss: IEEE, 1995, 12: 296-301.

[3] Huo Zhonghua, Chen Ying. Person re-identification based on multi-salience fusion[J]. Opto-Electronic Engineering, 2015, 42(9):41-47.

[4] Zheng W S, Gong S, Xiang T. Re-identification by relative distance comparison[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2013, 35(3): 653-668.

[5] Zhao R, Ouyang W, Wang X. Unsupervised salience learning for person re-identification[C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Portland: IEEE, 2013: 3586-3593.

[6] 范彩霞,陈亚军,曹磊,等. 基于视觉感知模型的行人再识别[J]. 计算机工程与应用,2016,52(6):156-161.
Fan Caixia, Chen Yajun, Cao Lei, et al. Person re-identification based on visual perceptual model[J]. Computer Engineering and Applications, 2016, 52(6): 156-161.

[7] 卢宏涛,张秦川. 深度卷积神经网络在计算机视觉中的应用研究综述[J]. 数据采集与处理,2016,31(1):1-17.
Lu Hongtao, Zhang Qinchuan. Application of deep convolutional neural network in computer vision[J]. Journal of Data Acquisition and Processing, 2016, 31(1): 1-17.

[8] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. Nature, 2015, 521(7553): 436-444.

[9] 曲建岭,杜辰飞,邸亚洲,等. 深度自动编码器的研究与展望[J]. 计算机与现代化,2014(8):128-134.
Qu Jiangling, Du Chenfei, Di Yazhou, et al. Research and prospect of deep auto-encoders[J]. Computer and Modernization, 2014(8):128-134.

[10] Masci J, Meier U, Ciresan D, et al. Stacked convolutional auto-encoders for hierarchical feature extraction[C]// Artificial Neural Networks and Machine Learning-ICANN 2011. Berlin Heidelberg: Springer, 2011: 52-59.

[11] Gray D, Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]// Proceeding of the 10th European Conference on Computer Vision. Marseille, France: Springer-Verlag, 2008, 5302: 262-275.

[12] Gray D,Tao H. Viewpoint invariant pedestrian recognition with an ensemble of localized features[C]// Computer Vision-ECCV 2008. Berlin Heidelberg: Springer,2008,5302: 262-275.

[13] Layne R, Hospedales T M, Gong S. Attributes-based re-identification[M]// Person Re-identification. London: Springer, 2014: 93-117.

[14] Umeda T, Sun Y, Irie G, et al. Attribute discovery for person re-identification[C] // International Conference on MultiMedia Modeling. North Miami: Springer International Publishing, 2016: 268-276.

[15] Ahmed E, Jones M, Marks T K. An improved deep learning architecture for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015: 3908-3916.

[16] Zheng W S, Gong S, Xiang T. Associating groups of people[C]// Proceedings of the 20th British Machine Vision Conference (BMVC). London: IEEE, 2009: 1-11.

[17] Layne R, Hospedales T M, Gong S, et al. Person re-identification by attributes[C]// Proceedings of the 23rd British Machine Vision Conference (BMVC). Guildford: IEEE, 2012, 2(3):8-18.

作者简介:

许方洁(1991-),女,硕士研究生,研究方向:机器学习及模式识别等,E-mail: daisyxfj@163.com。

成科扬(1982-),男,副教授,硕士生导师,研究方向:计算机应用技术和模式识别等。

张建明(1964-),男,教授,硕士生导师,研究方向:图象处理与模式识别、系统仿真、虚拟现实和计算机图形学等。

(编辑:王静)