

基于行人属性分级识别的行人再识别

陈鸿昶 吴彦丞* 李邵梅 高超

(国家数字交换系统工程技术研究中心 郑州 450002)

摘要: 为了提高行人再识别算法的识别效果, 该文提出一种基于注意力模型的行人属性分级识别神经网络模型, 相对于现有算法, 该模型有以下3大优点: 一是在网络的特征提取部分, 设计用于识别行人属性的注意力模型, 提取行人属性信息和显著性程度; 二是在网络的特征识别部分, 针对行人属性的显著性程度和包含的信息量大小, 利用注意力模型对属性进行分级识别; 三是分析属性之间的相关性, 根据上一级的识别结果, 调整下一级的识别策略, 从而提高小目标属性的识别准确率, 进而提高行人再识别的准确率。实验结果表明, 该文提出的模型相较于现有方法, 有效提高了行人再识别的首位准确率, 其中, Market1501数据集上, 首位准确率达到93.1%, 在DukeMTMC数据集上, 首位准确率达到81.7%。

关键词: 行人再识别; 注意力模型; 深度学习; 显著性; 属性分级

中图分类号: TP391.41

文献标识码: A

文章编号: 1009-5896(2019)09-2239-08

DOI: 10.11999/JEIT180740

Person Re-identification Based on Attribute Hierarchy Recognition

CHEN Hongchang WU Yancheng LI Shaomei GAO Chao

(China National Digital Switching System Engineering & Technological R&D Center, Zhengzhou 450002, China)

Abstract: In order to improve the accuracy rate of person re-identification, a pedestrian attribute hierarchy recognition neural network is proposed based on attention model. Compared with the existing algorithms, the model has the following three advantages. Firstly, the attention model is used in this paper to identify the pedestrian attributes, and to extract of pedestrian attribute information and degree of significance. Secondly, the attention model in used in this paper to classify the attributes according to the significance of the pedestrian attributes and the amount of information contained. Thirdly, this paper analyzes the correlation between attributes, and adjusts the next level identification strategy according to the recognition results of the upper level. It can improve the recognition accuracy of small target attributes, and the accuracy of pedestrian recognition is improved. The experimental results show that the proposed model can effectively improve the first accuracy rate (rank-1) of person re-identification compared with the existing methods. On the Market1501 dataset, the first accuracy rate is 93.1%, and the first accuracy rate is 81.7% on the DukeMTMC dataset.

Key words: Person re-identification; Attention model; Deep learning; Saliency; Hierarchy

1 引言

行人再识别一直是计算机视觉领域的研究热点, 但是由于行人外观易受光线、遮挡、视角和姿态的影响, 在不同的视频和图片中外观呈现出极大的差异性, 使得行人再识别成为一个极具挑战性的课题。随着深度学习在行人再识别领域的成功应用, 行人再识别的首位准确率相对于传统方法得到了大幅度提升, 但是由于大部分行人图片分辨率较

低, 而其中部分行人属性包含的信息量更是不足以保证正常的识别。因此, 有必要设计更为合理的网络结构, 提高信息量少的属性的识别准确率, 从而提高行人再识别的准确率。

目前的行人再识别算法大体可以分为3类, 分别是特征表示学习、距离度量学习方法和深度学习。其中特征表示学习基于特征表示向量, 计算行人之间的相似度, 得到行人再识别结果。文献[1]首先从图片提取行人前景, 基于行人前景区域的对称性和非对称性将前景划分成不同的区域, 对于每个区域, 提取带权重的颜色直方图等特征描述它们。文献[2]利用人体的颜色和结构信息, 基于空间直方图和区域协方差两种统计描述方法融合去识别

收稿日期: 2018-07-20; 改回日期: 2019-03-03; 网络出版: 2019-04-17

*通信作者: 吴彦丞 wuyc1994@163.com

基金项目: 国家自然科学基金(61601513)

Foundation Item: The National Natural Science Foundation of China (61601513)

行人。距离度量学习基于相关的测度学习算法，得到两张行人图片的相似度。其中文献[3]提出基于显著度融合的自适应分块方案，利用聚类的思想进行距离度量。文献[4]提出KISSME算法，该算法利用极大似然法判断两个行人向量是否相似。

深度学习方法是行人再识别领域近年来广泛使用的一种方法，文献[5]作为其中的代表，使用卷积神经网络(CNN)提取行人的特征，实现行人再识别。文献[6]将注意力模型结合进CNN网络中，提升了模型的特征提取能力。近年来，结合行人高层语义特征实现行人再识别逐渐成为主流方法，其中文献[7]不仅单独对每类行人属性进行识别，还将行人属性进行排列组合，然后联合识别多种属性和单个属性，文献[8]将行人属性和行人ID进行联合识别，文献[9]基于注意力模型，对行人进行分区域识别，得到了较高的识别准确率。

上述基于行人属性的深度学习方法有共同的特点，在识别行人属性过程中，没有充分考虑行人属性之间的相关性，以及属性包含信息量的差异性。但实际上，通过分析数据集，可以发现，有些属性之间具有强相关性，比如性别和下衣种类，其中穿裙子的样本大多数是女性；性别和头发长短也类似，长头发的样本大多数是女性。另外，不同行人属性在图片中代表的视觉面积不同，包含信息量具有较大差异，比如性别，年龄等属性，代表的是全局特征，包含信息量较大，头发、手提包、鞋子等属性，代表的是局部特征，包含信息量较少。而在网络识别中，包含信息量大的属性，更容易被准确识别。

针对现有研究存在的问题和上文分析，本文提出一种基于注意力模型的行人属性分级识别算法。该方法首先采用注意力模型对行人属性特征进行提取，得到特征表示向量和各行人属性的显著性程

度；其次基于显著性程度，对行人属性进行分级，等级高低作为后续属性识别顺序的依据；最后基于属性之间的相关性和高等级属性的识别结果，调整低等级属性的识别策略，从而提高低等级属性的识别效果。测试结果表明，本文算法在公共实验数据集上的实验结果，尤其是首位准确率，优于目前主流的行人再识别算法。

2 本文算法

在本节中，主要分为3个部分介绍基于注意力模型的行人属性分级识别算法，首先介绍算法整体网络架构以及注意力模型在算法中的应用；然后介绍属性分级的依据和结果；最后介绍如何根据属性分级和属性之间的相关性对识别策略进行调整。

2.1 网络整体结构与注意力模型

为了能更好识别不同种类的行人属性，提取到高鲁棒性的行人属性特征描述子，在文献[10]的基础上，本文设计了基于注意力模型的行人属性分级识别网络，具体网络结构见图1。本文网络主要由2个部分组成，以Conv10(卷积层)为分界线，前半部分为Vgg-16网络的前13层，后半部分为行人属性分级识别网络。下面详细介绍这2个方面的内容。

前半部分采用Vgg16网络作为特征提取网络，为了避免网络层数太深对后续注意力模型识别效果的影响，只选取Vgg16网络的前13层，其中包含了10层卷积层和3层池化层。网络的后半部分是行人属性分级识别网络，为了方便描述，先讨论batch=1的情况，即每次只输入一张尺寸为 $224 \times 224 \times 3$ 的图片，记为 $I=[224, 224, 3]$ 。在经过10层卷积和相应的池化计算后，将Conv10卷积层的输出记为 $O = \{Y_1, Y_2, \dots, Y_n\}$ ，代表输出的512个大小为[14, 14]的特征图(其中512为Vgg16网络默认值，后续的相关实验也表明该值为实验效果最佳

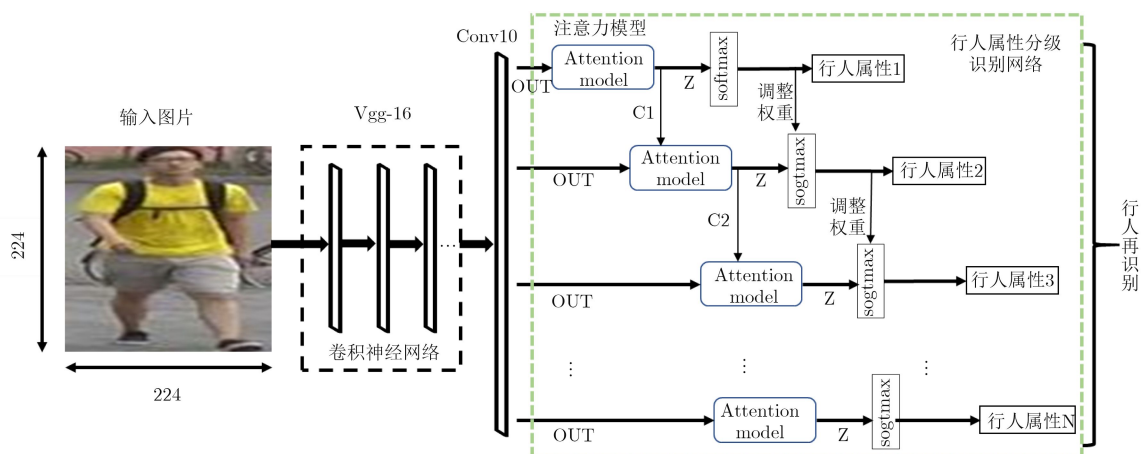


图1 网络整体结构

值), 其中 Y 代表每一个特征图, 大小为 $Y=[14, 14]$ 。将输出 O 作为行人属性分级识别网络的输入, 而该网络的核心结构就是注意力模型, 通过注意力模型判别人属性属于哪个区域, 注意力模型的内部结构见图2。

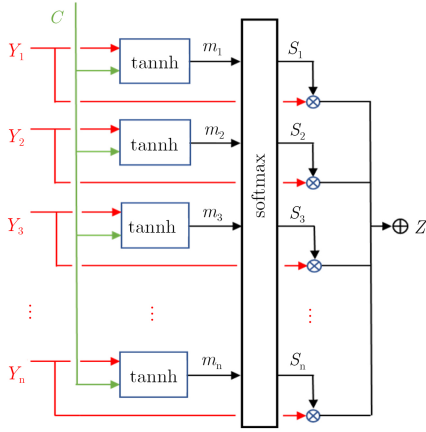


图 2 注意力模型网络结构

图2中 C 是上一级注意力模型识别的概率分布(初始为零分布), 作为输入, 和 Y 进行全连接, 然后经过激活函数 \tanh , 从图2中可以看出

$$m_i = \tanh(W_{cm}C + W_{ym}Y_i) \quad (1)$$

其中, W_{cm} 和 W_{ym} 是全连接时的参数, m_i 是经过激活函数后的输出。

m_i 经过 softmax 层, 得到行人属性属于各区域的概率 s_i , 最后的行人属性识别结果 z 则是由概率 s 和输入 Y 共同决定的, 计算方式如式(2)

$$z = \sum_{i=1}^n s_i y_i \quad (2)$$

而 Z 则成为下一级注意力模型的输入 C

$$C_{n+1} = Z_n \quad (3)$$

其中, n 表示注意力模型的级数。

2.2 属性分级

假设训练数据集中(以 Market1501 数据集为例, 见表1)包含 N 张行人图片, 分别属于 M 个不同的行人, 每张图片标注了 G 类属性, 包括性别, 头发长短, 是否带包, 上衣颜色等属性, 对于每一类属性, 其中包含了 K^g 种属性, 以上衣颜色为例, 其中包含黑色, 白色, 黄色等多种属性。将数据集用集合方式描述如式(4)

$$D = \{(x_i, \mathbf{l}_i^1, \mathbf{l}_i^2, \dots, \mathbf{l}_i^G)\}_{i=1}^N \quad (4)$$

其中, x_i 为第 i 张行人图片, 行人的第 g 类属性可以用向量 $\mathbf{l}_i^g = (\mathbf{l}_{i,1}^g, \mathbf{l}_{i,2}^g, \dots, \mathbf{l}_{i,K^g}^g)$ 表示, 每类属性中的第 k 种属性 $\mathbf{l}_{i,k}^g$ 都是二值向量表示, 即 $\mathbf{l}_{i,k}^g \in \{0, 1\}$, 如果行人存在该属性, 则 $\mathbf{l}_{i,k}^g = 1$, 反之则 $\mathbf{l}_{i,k}^g = 0$ 。

从表1中可以看出, 数据集中的行人属性标签被分为了 G 类, 相应地, 属性分级识别网络被划分为了 G 级。经过每一级注意力模型, 确定一个行人属性类所在区域, 再由 Softmax 层和损失函数确定每一级识别得到的结果。注意力模型的本质就是更易识别对结果影响最大的区域, 利用这一特性, 本文提出的分级识别网络首先识别出对行人识别结果影响最大的属性, 作为第1级, 而每一级注意力模型的输出 C 会和特征 Y 进行全连接以后作为下一级的输入, 将式(1)改写为

$$m_i = \tanh(W_{cm}C_i + W_{ym}Y_i) \quad (5)$$

在输入进第1级注意力模型时, 由于 C 为零分布, 对特征图 Y 不产生任何影响, 所以注意力模型输出的概率分布就是最能反映特征图真实情况的分布, 其中概率值最大的属性就是行人在网络中最显著的属性, 记为第1级属性。当特征图输入进第2级注意力模型时, 这时的 C 不再是零分布, 而是第

表 1 Market1501 数据集中的属性类别

属性类(G)	属性	数量(k)
Gender	male, female	2
Age	child, teenager, adult, old	4
Hair length	long, short	2
Length of lower-body clothing	long, short	2
Type of lower-body clothing	pants, dress	2
Wearing hat	yes, no	2
Carrying bag	yes, no	2
Carrying backpack	yes, no	2
Carrying handbag	yes, no	2
Color of upper-body clothing	black, white, red, purple, yellow, gray, blue, green	8
Color of lower-body clothing	black, white, pink, purple, yellow, gray, blue, green, brown	9

1级注意力模型输出的概率分布 Z , 这时由于 C 和特征图 Y 的全连接, 导致特征图中第1级属性的相关特征信息被抑制, 在输入第2级注意力模型以后, 由于最显著的属性被抑制, 得到的输出则是行人在网络中第2显著的属性。以此类推, 得到 G 类属性的层级划分。图3展示了行人属性分级网络中每一级的识别结果, 可以看出, 图例中, 网络首先识别的是更偏向于全局的、具有显著特征的属性, 随着级数的加深, 才会识别到一些面积较小的、不显著的属性。通过注意力分级机制, 可以得到属性的等级划分和各属性在样本中所在区域位置信息, 作为后续属性识别的基础。



图3 分级网络中每一级的识别结果

2.3 属性相关性及其识别策略调整

属性之间并不是独立的, 而是存在内部联系的, 通过分析数据集和人类经验知识可以发现, 一些属性之间存在强相关性, 比如女性、长头发、颜色鲜艳的衣服, 这些属性更容易同时出现, 男性、

短头发、运动服, 这些属性更容易同时出现, 在服饰的颜色上, 一般不会出现一些奇怪的配色, 比如上衣是红色, 下衣是绿色, 会比较符合人类正常审美。一些属性之间不存在太大的关联性, 比如年龄和性别, 戴帽子与否, 出现的概率比较独立。对本文使用的数据集(Market1501和DukeMTMC)进行属性间相关性分析(由于数据集在采集和人工标注属性信息时, 并没有统一, 导致数据集之间标签无论是数量上还是种类上都有较大差异, 所以每个数据集的属性间相关性分析需要单独做, 具有数据集特异性, 不便于统一计算), 选取具有代表性的属性绘制成图4, 其中的数值代表两者共现概率, 即当样本包含行属性(row)的基础上, 同时包含列属性(line)的概率, 计算公式为

$$P_k(\text{line/row}) = \frac{N_k(\text{line, row})}{N(\text{row})} \cdot \left\{ \sum_{k=1}^K P_k(\text{line/row}) = 1 \right\} \quad (6)$$

其中, $P_k(\text{line/row})$ 是条件概率, $N(\text{row})$ 是标注行属性样本的个数, $N_k(\text{line, row})$ 是标注了行属性的样本中同时标注了第 k 个列属性的样本个数。

在属性分级的基础上, 网络会根据分级结果和属性相关性, 分级识别属性。具体思想为根据上一级属性的识别结果, 基于属性之间的共现概率, 调整不同共现概率大小的属性种类在下一级识别时的权重, 例如在识别性别为男性时, 后续识别过程中, 会抑制识别头发为长头发, 下衣种类为裙子的识别结果, 提高头发为短头发, 下衣种类为裤子的识别结果的权重, 从而减少误识的出现。在识别第1级行人属性时, 本文网络使用softmax层对属性进行分类识别, 对于属性分类的结果, 本文采用交叉信息熵损失函数进行计算, 可以表示为

	female	male	short hair	teenager	adult	short sleeve	long lower	short lower	dress	pants	backpack	handbag	downblack	upwhite
female	1.00	0.00	0.25	0.79	0.18	0.91	0.33	0.68	0.34	0.66	0.22	0.13	0.34	0.35
male	0.00	1.00	0.99	0.73	0.23	0.97	0.44	0.56	0.00	1.00	0.30	0.10	0.43	0.27
short hair	0.16	0.84	1.00	0.71	0.25	0.97	0.43	0.57	0.04	0.96	0.28	0.11	0.41	0.27
teenager	0.45	0.55	0.63	1.00	0.00	0.95	0.30	0.70	0.15	0.85	0.34	0.09	0.36	0.31
adult	0.37	0.63	0.80	0.00	1.00	0.96	0.71	0.29	0.12	0.88	0.05	0.23	0.53	0.28
short sleeve	0.41	0.59	0.69	0.76	0.22	1.00	0.39	0.61	0.14	0.86	0.27	0.12	0.39	0.30
long lower	0.35	0.65	0.74	0.57	0.39	0.94	1.00	0.00	0.05	0.95	0.21	0.16	0.46	0.35
short lower	0.47	0.53	0.63	0.88	0.10	0.96	0.00	1.00	0.21	0.79	0.30	0.09	0.34	0.28
dress	1.00	0.00	0.17	0.78	0.17	0.90	0.14	0.86	1.00	0.00	0.20	0.12	0.25	0.32
pants	0.33	0.67	0.76	0.75	0.22	0.96	0.44	0.56	0.00	1.00	0.28	0.11	0.41	0.30
backpack	0.36	0.64	0.72	0.96	0.04	0.95	0.31	0.69	0.11	0.89	1.00	0.00	0.36	0.26
handbag	0.50	0.50	0.66	0.57	0.42	0.98	0.53	0.47	0.15	0.85	0.00	1.00	0.37	0.29
downblack	0.37	0.63	0.71	0.69	0.29	0.95	0.46	0.54	0.10	0.90	0.25	0.11	1.00	0.31
upwhite	0.50	0.50	0.60	0.79	0.20	0.94	0.45	0.55	0.15	0.85	0.23	0.11	0.40	1.00

(a) Market501数据集属性间共现矩阵

	female	male	bag	boots	backpack	hat	dark	long	short	downblue	upblack	upwhite	upgray	upblue
female	1.00	0.00	0.24	0.55	0.60	0.09	0.83	0.26	0.74	0.28	0.60	0.11	0.09	0.06
male	0.00	1.00	0.11	0.06	0.69	0.19	0.87	0.07	0.93	0.35	0.62	0.05	0.12	0.12
bag	0.63	0.37	1.00	0.34	0.00	0.16	0.88	0.31	0.68	0.27	0.60	0.09	0.08	0.10
boots	0.88	0.12	0.21	1.00	0.65	0.09	0.97	0.24	0.75	0.24	0.63	0.09	0.08	0.07
backpack	0.40	0.60	0.00	0.27	1.00	0.13	0.84	0.10	0.90	0.34	0.63	0.07	0.11	0.09
hat	0.28	0.72	0.18	0.16	0.55	1.00	0.79	0.16	0.84	0.32	0.48	0.11	0.13	0.14
dark	0.43	0.57	0.17	0.31	0.64	0.14	1.00	0.15	0.85	0.31	0.61	0.07	0.11	0.10
long	0.74	0.26	0.35	0.43	0.44	0.16	0.84	1.00	0.00	0.24	0.66	0.07	0.10	0.03
short	0.38	0.62	0.13	0.24	0.69	0.15	0.85	0.00	1.00	0.33	0.61	0.08	0.11	0.11
downblue	0.38	0.62	0.14	0.21	0.69	0.15	0.82	0.12	0.88	1.00	0.60	0.09	0.07	0.10
upblack	0.43	0.57	0.16	0.28	0.67	0.12	0.84	0.16	0.83	0.31	1.00	0.00	0.00	0.00
upwhite	0.61	0.39	0.20	0.33	0.56	0.20	0.76	0.15	0.85	0.37	0.00	1.00	0.00	0.00
upgray	0.37	0.63	0.13	0.20	0.63	0.17	0.88	0.14	0.86	0.21	0.00	0.00	1.00	0.00
upblue	0.30	0.70	0.18	0.21	0.65	0.23	0.89	0.05	0.95	0.35	0.00	0.00	0.00	1.00

(b) DukeMTMC数据集属性间共现矩阵

图4 数据集部分属性间相关性及其共现概率

$$L^g = -\frac{1}{N^g} \frac{1}{K^g} \sum_{i=1}^{N^g} \sum_{k=1}^{K^g} \left[l_{i,k}^g \lg \left(p \left(\hat{l}_{i,k}^g \right) \right) + (1 - l_{i,k}^g) \lg \left(1 - p \left(\hat{l}_{i,k}^g \right) \right) \right] \quad (7)$$

而对于下一级属性，需要基于上一级属性的识别结果，赋予属性类中各子类不同“偏好”权重，主要对softmax层计算公式进行变换，将其改进为

$$\left. \begin{aligned} p \left(\hat{l}_{i,k}^g \right) &= \omega \frac{\exp \left(\hat{l}_{i,k}^g \right)}{\sum_{j=1}^{K^g} \hat{l}_{i,j}^g} \\ \omega &= \exp \left(P_k(\text{line/row}) \right) \end{aligned} \right\} \quad (8)$$

以上都是针对单个属性识别进行的损失计算，对于网络整体而言，需要一个联合损失函数，保证网络的迭代和收敛，网络整体损失函数可以表示为

$$\left. \begin{aligned} \text{Loss} &= \frac{1}{G} \sum_{g=1}^G L^g \\ L^g &= -\frac{1}{N^g} \frac{1}{K^g} \sum_{i=1}^{N^g} \sum_{k=1}^{K^g} \left[l_{i,k}^g \lg \left(p \left(\hat{l}_{i,k}^g \right) \right) + (1 - l_{i,k}^g) \lg \left(1 - p \left(\hat{l}_{i,k}^g \right) \right) \right] \end{aligned} \right\} \quad (9)$$

在网络损失函数优化时，采用批量梯度下降的方法，即每次计算一个批处理样本的损失，这使得网络在最优化过程中能降低计算量，而且由于损失函数是凸函数，所以可以保证收敛到一个较好的全局最优解。

2.4 行人特征向量度量方法

本文的目标是实现行人再识别，而不是单纯的属性识别，所以从网络中softmax层之前，提取出每个属性的特征向量，合成为行人特征向量，具体做法如下：以Market1501数据集为例，本文将属性分为了 G 类，则相应地假设行人特征向量为 $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_G]$ ，其中， $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_G$ 分别表示一类行人属性，该顺序是按照表1的属性出现顺序固定下来的，当本文网络识别到对应的属性时，则在行人特征向量相应的部分添加向量值，没有识别到相应的属性时，则该部分为零向量。随后根据得到行人特征向量，采用交叉视角的二次判别分析法(XQDA)^[11]进行向量之间距离的度量，该方法是在KISSME算法和贝叶斯方法基础上提出的。该方法用高斯模型分别拟合类内和类间样本特征的差值分布。根据两个高斯分布的对数似然比推导出马氏距离，以 \sum 表示高斯分布，则两个样本之间的距离为

$$d(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \left(\sum_I^{-1} - \sum_E^{-1} \right) (\mathbf{x}_i - \mathbf{x}_j) \quad (10)$$

在测试时，计算待识别样本和查询集所有样本之间的距离，其中每两者之间的距离都包含两个距离值，第1个距离值是只计算两者都是非零向量部分之间的距离值，也就是只计算两者都识别到的属性之间的距离，记为 d_1 ，第2个距离值是计算完整行人向量之间的距离值，记为 d_2 。最后对所有计算得到的距离值进行排序，首先对 d_1 进行排序，选取距离值最小的10个查询集样本，随后比较这些样本的与待识别样本之间的 d_2 值，选取距离最小的样本作为预测识别样本，得到首位准确率的样本，也就是行人再识别结果。

3 实验结果与分析

本文实验是基于深度学习框架Matconvnet实现的，实验平台是配备64 GB内存的Intel core i7处理器和24 GB显存的Nvidia TITAN X显卡的GPU工作站。

3.1 数据集和评价指标

本文主要基于两个具有行人属性标注的行人再识别数据集进行实验，分别是Market1501, Duke-MTMC数据集。为了与已有算法公正比较，实验中，采用先前工作普遍采用的评价框架。主要有以下两种评价指标：一种是使用累积匹配特性(Cumulative Match Characteristic, CMC)曲线评价算法性能，另一种是采用平均准确率(mean Average Precision, mAP)评价算法性能。

3.2 网络参数与结构设置

本文网络参数是基于注意力模型^[10]原参数的基础上调整优化得到的，并将该模型作为基础网络(baseline)进行实验结果的对比，在训练过程中，将数据集图片尺寸统一调整为 224×224 ，作为网络输入，设置批处理尺寸(batch size)大小为64，循环次数(epochs)为80，学习率初始值为0.001，在随后每经过1000次迭代，学习率减小5%。下面就网络训练的相关设置进行详细描述。

迭代次数 如图5(a)所示，其中曲线代表了当网络迭代不同次数时，在不同的数据集上，网络的首位准确率变化情况。每迭代1000次测试1次网络性能，随着迭代次数达到14000次左右，网络性能基本稳定，达到最大首位准确率，所以将网络的循环次数设置为80。

属性分级 如图5(b)所示，以不加改进的注意力模型作为基础网络，在此基础上，使用两种属性分级策略，一是按照人类主观认知，根据属性在样本图片中所占的面积事先对各属性进行分级(以Market1501数据集为例，属性等级划分为：性别>年龄>头发长短>上衣种类>下衣种类>上衣颜色

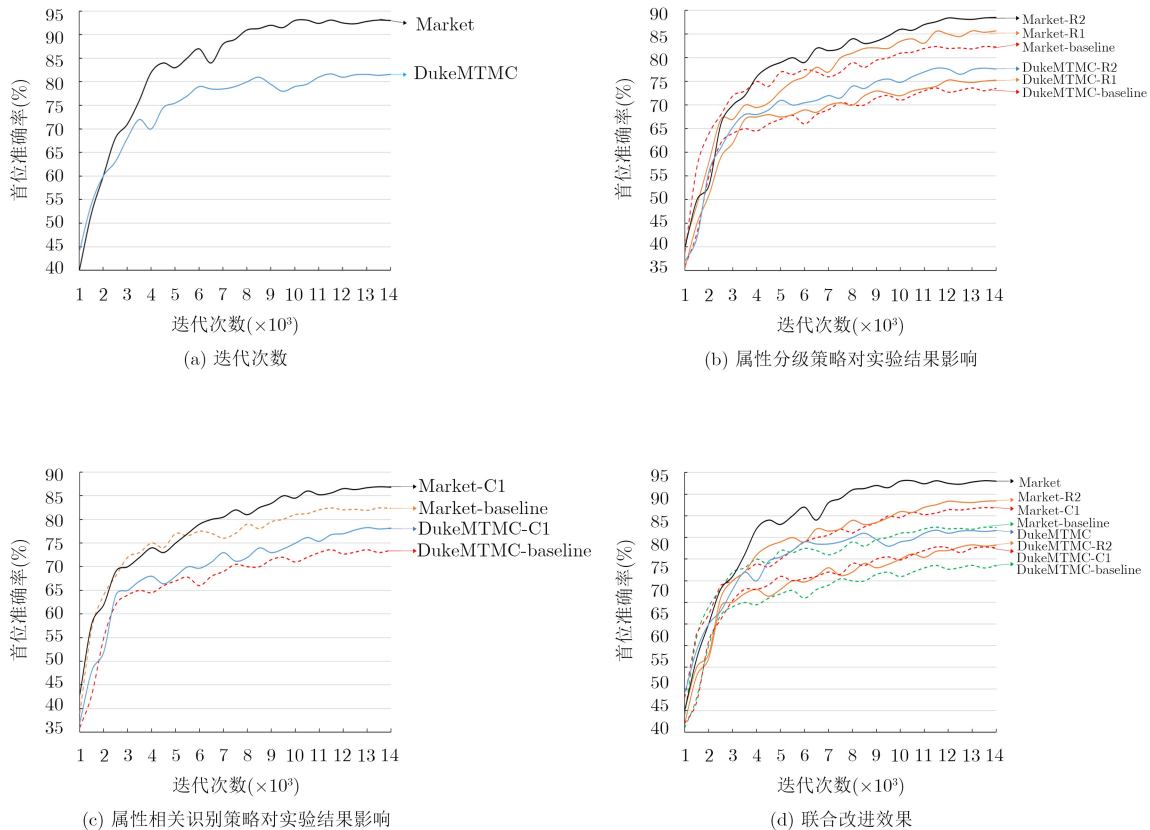


图5 网络参数及改进效果对比

>下衣颜色>下衣长短>背包有无>手提包有无>帽子有无), 在图例中记为R1; 二是按照本文提出的, 基于注意力模型特性而得到的自适应分级, 在图例中记为R2。基于以上两种策略分别进行实验, 可以发现, 相对于直接使用注意力模型进行行人再识别, 两种属性分级策略在首位准确率上都有提升作用, 在Market1501数据集上首位准确率分别提升了3.3%, 6.0%, 在DukeMTMC数据集上首位准确率分别提升了1.7%, 4.2%。可见本文提出的分级策略要比按照人类主观认知进行分级得到的实验效果要好。

属性相关性 如图5(c)所示, 同样地, 以不加改进的注意力模型作为基础网络, 在此基础上, 添加基于属性相关性的识别调整策略, 分别在不同的数据集上进行实验, 在图例中记为C1。可以从图中实验结果发现, 相较于基础网络, 基于属性相关性对识别策略进行调整后的网络, 在Market1501, DukeMTMC数据集上首位准确率分别提升了4.5%,

4.7%。说明在网络中添加属性相关性信息可以提高网络的识别能力。

联合改进效果 如图5(d)所示, 分别对比基础网络, 单独添加属性分级策略, 单独添加属性相关性调整策略和联合两者改进策略(本文算法)这4种算法。可以发现, 相对于对网络进行单一的改进, 本文提出的联合改进策略对网络的性能提升最佳, 首位准确率也是较单一改进有所提高。

3.3 各属性的识别精度

本文基于公开行人属性数据集Market1501和DukeMTMC, 分别进行了行人属性的识别实验, 识别准确率如表2, 表3所示。选取不加改进的注意力模型作为基础网络, 进行实验结果的比对。从属性识别的平均准确率来看, 在这两个数据集上, 本文的属性识别平均准确率相较于基础网络都有了较大的提升, 平均准确率分别提升了7.41%和6.83%, 可见本文算法在基础网络的基础上有较大的改进。

表2 Market1501数据集各属性识别准确率(%)

行人属性	gender	age	hair	L.slv	L.low	S.cloth	B.pack	H.bag	bag	hat	C.up	C.low	mean
基础网络	82.18	85.32	80.12	92.48	71.58	85.67	79.57	81.54	79.66	70.56	91.23	87.81	82.31
本文算法	90.27	88.15	91.54	93.55	87.25	90.48	89.77	87.65	84.67	87.39	92.44	93.48	89.72

表3 DukeMTMC数据集各属性识别准确率(%)

行人属性	gender	hat	boots	L.up	B.pack	H.bag	bag	C.shoes	C.up	C.low	mean
基础网络	82.47	75.48	76.14	73.58	71.58	69.42	78.31	68.54	62.17	51.24	70.89
本文算法	83.59	87.24	84.56	76.33	77.11	75.32	83.78	72.19	74.88	62.18	77.72

3.4 行人再识别结果

本文基于两个通用行人属性数据集,进行了行人再识别实验,实验结果如表4,表5所示。其中“基础网络”表示基于不加改进的注意力模型得到的行人再识别结果,“基础网络-R1”表示在注意力模型的基础上单独添加基于人类主观认知的分级策略得到的结果,“基础网络-R2”表示在注意力模型的基础上单独添加基于模型自适应分级的分级策略得到的结果,“基础网络-C1”表示在注意力模型的基础上单独添加属性相关性识别调整策略得到的结果。

表4 Market1501数据集行人再识别结果(%)

方法	Rank-1	mAP
XQDA ^[11]	43.8	22.2
SCS ^[12]	51.9	26.3
DNS ^[13]	61.0	35.6
G-SCNN ^[14]	65.8	39.5
MSCAN ^[15]	80.3	57.5
PDC ^[16]	84.1	63.4
JLML ^[17]	85.1	65.5
HA-CNN ^[8]	91.2	75.7
基础网络	82.4	61.2
基础网络-R1	85.7	66.7
基础网络-R2	88.4	70.3
基础网络-C1	86.9	68.5
本文算法	93.1	76.2

表5 DukeMTMC数据集行人再识别结果(%)

方法	Rank-1	mAP
BoW+KISSME ^[18]	25.1	12.2
LOMO+XQDA ^[11]	30.8	17.0
ResNet50 ^[19]	65.2	45.0
ResNet50+LSRO ^[20]	67.7	47.1
JLML ^[17]	73.3	56.4
HA-CNN ^[8]	80.5	63.8
基础网络	73.6	55.7
基础网络-R1	75.3	57.4
基础网络-R2	77.8	60.8
基础网络-C1	78.3	61.2
本文算法	81.7	65.9

3.4.1 Market1501数据集

首先,从表4中可以看出,相对于基础网络,在进行单一改进后(添加了属性分级策略和属性相关性识别策略),网络的首位准确率分别提升了6.0%,4.5%,这两种策略都充分挖掘了属性之间的内在联系信息,提高了属性的识别效果。相应地,平均准确率也分别提升了9.1%,7.3%。与本文完整算法相比,单一改进识别策略得到的首位准确率有所下降,由此可见,将两种分类方式进行结合,可以得到更好的实验效果。

3.4.2 DukeMTMC数据集

同样地,从表5中可以发现,相对于基础网络,单一添加属性分级策略和属性相关性识别策略后,网络的首位准确率分别提升了4.2%,4.7%,相应地,平均准确率也分别提升了5.1%,5.5%。类似于Market1501数据集的实验结果,单一改进策略并没有联合使用得到的结果好。其次,针对DukeMTMC数据集,HA-CNN网络在该数据集上达到了80.5%的首位准确率和63.8%的平均准确率。本文在现有方法的基础上进一步提高实验效果,在该数据集上达到了81.7%的首位准确率和65.9%的平均准确率。

4 结束语

随着深度学习在行人再识别领域的成功应用,从最初的卷积神经网络到现在的注意力模型、对抗学习,行人再识别的精度也是越来越高。本文在现有工作的基础上,对注意力模型进行了改进,提出了属性分级识别策略,进一步提高了行人再识别的首位准确率。本文不仅从理论上阐述了分级识别策略的合理性,在两个公开数据集上的实验结果也证明了本文策略的有效性。未来将会在网络架构中添加属性异质性信息提升识别效果。

参考文献

- [1] FARENZENA M, BAZZANI L, PERINA A, *et al.* Person re-identification by symmetry-driven accumulation of local features[C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 2010: 2360–2367.
- [2] 曾明勇, 吴泽民, 田畅, 等. 基于外观统计特征融合的人体目标再识别[J]. 电子与信息学报, 2014, 36(8): 1844–1851. doi: 10.3724/SP.J.1146.2013.01389.

- ZENG Mingyong, WU Zemin, TIAN Chang, *et al.* Fusing appearance statistical features for person re-identification[J]. *Journal of Electronics & Information Technology*, 2014, 36(8): 1844–1851. doi: [10.3724/SP.J.1146.2013.01389](https://doi.org/10.3724/SP.J.1146.2013.01389).
- [3] 陈鸿昶, 陈雷, 李邵梅, 等. 基于显著度融合的自适应分块行人再识别[J]. *电子与信息学报*, 2017, 39(11): 2652–2660. doi: [10.11999/JEIT170162](https://doi.org/10.11999/JEIT170162).
- CHEN Hongchang, CHEN Lei, LI Shaomei, *et al.* Person re-identification of adaptive blocks based on saliency fusion[J]. *Journal of Electronics & Information Technology*, 2017, 39(11): 2652–2660. doi: [10.11999/JEIT170162](https://doi.org/10.11999/JEIT170162).
- [4] KÖSTINGER M, HIRZER M, WOHLHART P, *et al.* Large scale metric learning from equivalence constraints[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, USA, 2012: 2288–2295.
- [5] LI Wei, ZHAO Rui, XIAO Tong, *et al.* DeepReID: Deep filter pairing neural network for person re-identification[C]. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 152–159.
- [6] LIU Hao, FENG Jiashi, QI Meibin, *et al.* End-to-end comparative attention networks for person re-identification[J]. *IEEE Transactions on Image Processing*, 2017, 26(7): 3492–3506. doi: [10.1109/TIP.2017.2700762](https://doi.org/10.1109/TIP.2017.2700762).
- [7] MATSUKAWA T and SUZUKI E. Person re-identification using CNN features learned from combination of attributes[C]. The 23rd International Conference on Pattern Recognition, Cancun, Mexico, 2016: 2428–2433.
- [8] LI Wei, ZHU Xiatian, and GONG Shaogang. Harmonious attention network for person re-identification[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018, 2285–2294.
- [9] LIN Yutian, ZHENG Liang, ZHENG Zhedong, *et al.* Improving person re-identification by attribute and identity learning[EB/OL]. <http://arxiv.org/abs/1703.07220>, 2017.
- [10] XU K, BA J, KIROUS R, *et al.* Show, attend and tell: Neural image caption generation with visual attention[C]. 2015 International Conference on Machine Learning, New York, USA, 2015: 2048–2057.
- [11] LIAO Shengcai, HU Yang, ZHU Xiangyu, *et al.* Person re-identification by local maximal occurrence representation and metric learning[C]. 2015 IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 2197–2206.
- [12] CHEN Dapeng, YUAN Zejian, CHEN Badong, *et al.* Similarity learning with spatial constraints for person re-identification[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1268–1277.
- [13] ZHANG Li, XIANG Tao, and GONG Shaogang. Learning a discriminative null space for person re-identification[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1239–1248.
- [14] VARIOR R R, HALOI M, and WANG Gang. Gated Siamese convolutional neural network architecture for human re-identification[C]. The 14th European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 791–808.
- [15] LI Dangwei, CHEN Xiaotang, ZHANG Zhang, *et al.* Learning deep context-aware features over body and latent parts for person re-identification[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7398–7407.
- [16] SU Chi, LI Jianing, ZHANG Shiliang, *et al.* Pose-driven deep convolutional model for person re-identification[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3980–3989.
- [17] LI Wei, ZHU Xiatian, and GONG Shaogang. Person re-identification by deep joint learning of multi-loss classification[C]. The 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia, 2017: 2194–2200.
- [18] WANG Hanxiao, GONG Shaogang, and XIANG Tao. Highly efficient regression for scalable person re-identification[EB/OL]. <http://arxiv.org/abs/1612.01341>, 2016.
- [19] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778.
- [20] ZHENG Zhedong, ZHENG Liang, and YANG Yi. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3774–3782.
- 陈鸿昶: 男, 1964年生, 教授, 博士生导师, 研究方向为通信与信息系统、计算机视觉。
- 吴彦丞: 男, 1994年生, 硕士生, 研究方向为计算机视觉。
- 李邵梅: 女, 1982年生, 博士, 讲师, 研究方向为通信与信息系统、计算机视觉。
- 高超: 男, 1982年生, 博士, 讲师, 研究方向为通信与信息系统、计算机视觉。