Journal of Computer Applications
计算机应用,2020,40(9):2493 - 2498
ISSN 1001-9081
CODEN JYIIDU
2020-09-10
http://www.joca.cn

文章编号:1001-9081(2020)09-2493-06
DOI:10.11772/j.issn.1001-9081.2020010006

基于生成对抗网络联合时空模型的行人重识别方法

邱耀儒,孙为军*,黄永慧,唐瑜祺,张浩川,吴俊鹏
(广东工业大学 自动化学院,广州 510006)
(*通信作者电子邮箱 gdutswj@gdut.edu.cn)

摘　要:跨摄像头的行人跟踪是智慧城市和智能安防的技术难题,而行人重识别是跨摄像头行人跟踪中最关键的技术。针对因领域偏差而导致行人重识别算法跨场景应用时准确率大幅下降的问题,提出一种基于生成对抗网络(GAN)联合时空模型的方法(STUGAN)。首先,通过引入GAN生成目标场景的训练样本,以增强识别模型的稳定性;然后,利用时空特征构建目标场景样本的时空模型,以筛选低概率匹配样本;最后,融合识别模型与时空模型来进行行人的重识别。在该领域典型的公开数据集Market-1501和DukeMTMC-reID上与词袋模型(BoW)、先进的无监督学习(PUL)、无监督多任务字典学习(UMDL)等先进无监督算法进行实验对比。实验结果分析表明,所提方法在Market-1501数据集上的rank-1、rank-5及rank-10指标的准确率达到66.4%、78.9%及84.7%,比对比算法最好结果分别高5.7,5.0和4.4个百分点;平均精度均值(mAP)仅低于保持主题一致性的循环生成网络(SPGAN)。

关键词:行人跟踪;行人重识别;跨场景应用;生成对抗网络;时空模型
中图分类号:TP391.41　　**文献标志码**:A

Person re-identification method based on GAN uniting with spatial-temporal pattern

QIU Yaoru, SUN Weijun*, HUANG Yonghui, TANG Yuqi, ZHANG Haochuan, WU Junpeng
(*School of Automation, Guangdong University of Technology, Guangzhou Guangdong* 510006, *China*)

Abstract: Tracking of the person crossing the cameras is a technical challenge for smart city and intelligent security. And person re-identification is the most important technology for cross-camera person tracking. Due to the domain bias, applying person re-identification algorithms for cross-scenario application leads to the dramatic accuracy reduction. To address this challenge, a method based on Generative Adversarial Network (GAN) Uniting with Spatial-Temporal pattern (STUGAN) was proposed. First, training samples of the target scenario generated by the GAN were introduced to enhance the stability of the recognition model. Second, the spatio-temporal features were used to construct the spatio-temporal pattern of the target scenario, so as to screen low-probability matching samples. Finally, the recognition model and the spatio-temporal pattern were combined to realize the person re-identification task. On classic datasets of this field named Market-1501 and DukeMTMC-reID, the proposed method was compared with BoW (Bag-of-Words), PUL (Progressive Unsupervised Learning), UMDL (Unsupervised Multi-task Dictionary Learning) and other advanced unsupervised algorithms. The experimental results show that the proposed method achieves 66.4%, 78.9% and 84.7% recognition accuracy for rank-1, rank-5 and rank-10 indicators on the Market-1501 dataset respectively, which are 5.7, 5.0 and 4.4 percentage points higher than the best results of the comparison algorithm, respectively; and the mean Average Precision (mAP) higher than the comparison algorithms except Similarity Preserving cycle-consistent Generative Adversarial Network (SPGAN).

Key words: person tracking; person re-identification; cross-scenario application; Generative Adversarial Network (GAN); spatial-temporal pattern

0　引言

目前中国正在大量地建设智慧城市,智能安防是实现智慧城市的重要基础,并且智能安防管理中每天产生的视频数据占据城市数据的很大比重。面对大量的视频数据,仅依靠传统的人工手段无法做到快速、高效的分析和处理,因此需要将计算机视觉以及最新的计算机技术应用于其中,以实现高效、准确的视频分析。

行人重识别(Person re-identification)[1]作为视频分析研究的关键组成部分,目的是对出现在监控区域内的某个目标行人,在监控网络的其他摄像头监控视频中找到目标并标识出来。配合行人检测与跟踪技术可以实现对监控网络范围内的

收稿日期:2020-01-16;**修回日期**:2020-05-08;**录用日期**:2020-05-08。
基金项目:国家重点研发计划项目(2018YFB1802400);广东省重点领域研发计划项目(2019B010118001, 2019B010121001)。
作者简介:邱耀儒(1993—),男,广东揭阳人,硕士研究生,主要研究方向:深度学习、图像识别;　孙为军(1975—),男,安徽当涂人,讲师,博士,CCF会员,主要研究方向:智能信息处理;　黄永慧(1975—),女,湖北麻城人,讲师,博士,主要研究方向:图像处理、模式识别;　唐瑜祺(1995—),女,湖南邵阳人,硕士研究生,主要研究方向:深度学习、图像识别;　张浩川(1994—),男,江西抚州人,硕士研究生,主要研究方向:深度学习、计算机视觉;　吴俊鹏(1994—),男,广东广州人,硕士研究生,主要研究方向:机器学习、模式识别。

www.joca.cn

目标行人进行跨摄像头区域的持续跟踪。当目标人物出现在某一监控摄像头的监控范围内时，摄像头拍摄的目标行人图像会与其他摄像头拍摄的行人图像进行一一匹配，寻找出同属于目标行人的图像，从而进行跨摄像机的搜索与跟踪。行人重识别技术的应用可以减少大量视频监控人员的工作量，实现监控视频中目标行人的准确分析，有效地预防和打击犯罪行为。经过近10年的发展，研究人员提出了大量的行人重识别模型，但是实际监控环境的复杂多变，导致行人重识别技术依然存在诸多问题没有解决。影响行人重识别问题主要有图像像素分辨率低、行人姿态角度多样化、光照变化等因素。如图1所示，同一行人在不同摄像头拍摄下的图像呈现诸多差异。图1(a)~(e)为同一行人由不同摄像头拍摄的图像。图1(a)与图1(f)为同一摄像头拍摄下的不同行人，图1(b)与图1(g)相同，以此类推。



图1 不同摄像头拍摄的行人样本

Fig. 1 Person samples taken by different cameras

目前行人重识别模型主要分为行人特征表达模型和行人相似性判别模型等类型。特征表达方法集中于行人外观信息特征的提取和表达上，使得不同行人的特征具有判别性。最开始的行人方法研究是通过手工设计的特征对行人信息进行描述，Ojala等[2]和Swain等[3]提出基于颜色或纹理信息的特征表示方法。在此基础上，基于集成特征学习的方法[4]被提出并取代了手工设计特征的方法成为经典的行人特征表达方法。国内外研究人员从底层特征方向入手，相继提出了基于跨领域不变颜色特征方法[5]、基于局部对齐的特征变换方法[6]、基于局部最大出现频率的表达模型[7]等行人特征表达方法。近年来，随着深度学习的应用，越来越多的深度学习方法被用来提取行人图像的抽象特征。Ahemd等[8]、Chen等[9]和陈兵等[10]利用卷积神经网络提取行人图像的全局抽象特征，所提取的特征比手工设计的特征具有更好的判别性。相似性判别方法的主要思想是对行人训练样本进行相似性判别学习，从样本中学习最优的相似性度量指标。文献[11-15]提出有监督的行人相似性度量判别方法，使得同一行人的距离最小而不同行人的距离最大。实际监控中存在大量无标签信息的数据，因此Peng等[16]和Yu等[17]提出基于无监督学习的相似性度量模型。近年来，部分国内外学者开始对行人重识别中的其他特定问题展开研究。从数据角度出发，部分学者提出迁移学习模型[18]，利用其他场景中已有标注的行人数据进行模型训练，然后将模型应用于目标场景。Zhu等[19]提出了从图片风格层面进行迁移的循环生成对抗网络(Cycle-consistent Generative Adversarial Network，CycleGAN)，Deng等[20]在此基础上加入行人限制损失，提出保持图像生成前后主体一致性的循环生成对抗网络(Similarity Preserving cycle-consistent Generative Adversarial Network，SPGAN)。除了利用视觉信息进行图像匹配，Huang等[21]和Martinel等[22]提出多模态学习模型，通过利用行人图像的时间及空间信息辅助模型判断。Lv等[23]提出了视觉融合时空模型(Transfer Learning of Spatio-temporal Pattern，TFusion)算法，通过贝叶斯策略融合时空模型和视觉特征模型进行行人图像的判别。

现有的研究虽取得一定的成效，但是大部分的算法属于有监督学习的方法，这类方法在单场景应用中能够获得相对较高的准确率。但是当算法需要部署于新的应用场景时，需要耗费大量的成本构建对应场景的样本标签，这类算法应用比较局限。其中的无监督方法虽然能够在多个不同场景中应用，但是由于缺少标签信息训练，算法准确率偏低，达不到实际应用要求。因此本文提出一种基于生成对抗网络联合时空模型的方法，通过引入生成对抗网络生成目标场景的样本训练特征提取模型，使得模型适应于目标场景样本的识别。在此基础上利用目标场景的数据生成时空模型筛选低概率匹配样本，提高模型在目标应用场景的性能表现。依据模型特性，模型命名为基于生成对抗网络联合时空模型(GAN Uniting with Spatio-Temporal Pattern，STUGAN)算法。通过在两个标准的行人重识别数据集Market-1501[24]和DukeMTMC-reID[25]进行实验，验证了算法的有效性。

本文的主要贡献如下：

1)利用图像分类器有效联合生成对抗网络和时空模型，生成对抗网络的生成样本用于提升图像分类器在目标场景的识别准确率，高水准的图像分类器构建更真实的时空模型。

2)训练过程不需要目标场景的数据标签，使得模型能够快速应用于不同的目标场景，有效减少实际项目工程中构建数据标签的时间与人力资源。

1 相关模型

1.1 SPGAN算法

SPGAN算法是在CycleGAN的基础上加入行人限制损失函数，能够更好地保留行人图像的主体信息。SPGAN需要训练两对生成器与鉴别器，分别为$\{G, D_T\}$和$\{F, D_S\}$。生成器$G(\cdot)$的目标是生成能够欺骗鉴别器D_T的样本，而鉴别器的任务是辨别样本是否为生成样本，两者的对抗损失函数如下：

$$L_{T\text{adv}}(G, D_T, p_x, p_y) = E_{y\sim p_y}[(D_T(y)-1)^2] + E_{x\sim p_x}[D_T(G(x))^2] \tag{1}$$

其中：p_x和p_y表示源域和目标域的样本分布，样本$x \in S, y \in T$(S表示源数据集，T表示目标数据集)。

生成器$F(\cdot)$和D_S的任务与$G(\cdot)$和D_T的任务相同，但数据样本的生成方向相反，两者的对抗损失函数如下：

$$L_{S\text{adv}}(G, D_S, p_x, p_y) = E_{x\sim p_x}[(D_S(x)-1)^2] + E_{y\sim p_y}[D_S(F(y))^2] \tag{2}$$

通常情况，生成对抗网络通常需要对齐的样本对进行训练，而CycleGAN使用了循环自监督模型$F(G(S)) \approx S$，无需进行样本标注。循环损失约束表示如下：

$$L_{\text{cyc}}(G, F) = E_{x\sim p_x}\left[\left\|F(G(x))-x\right\|_1\right] + E_{y\sim p_y}\left[\left\|G(F(y)-y)\right\|_1\right] \tag{3}$$

CycleGAN中使用了一致性损失函数减小生成样本的色彩偏差，其损失函数如下：

www.joca.cn

$$L_{\text{ide}}(G,F,p_x,p_y)=E_{x\sim p_x}\left\|F(x)-x\right\|_1+E_{y\sim p_y}\left\|G(y)-y\right\|_1 \quad (4)$$

式(1)~(4)由CycleGAN提出,SPGAN在此基础上加入孪生网络进行生成样本间的比较,目的是保留生成图像的行人主体信息,其损失函数如下:

$$L_{\text{con}}(i,\boldsymbol{x}_1,\boldsymbol{x}_2)=(1-i)\{\max(0,m-d)\}^2+id^2 \quad (5)$$

其中:$\boldsymbol{x}_1$和$\boldsymbol{x}_2$是经过孪生网络提取后的向量特征样本对,d表示$\boldsymbol{x}_1$和$\boldsymbol{x}_2$之间的欧氏距离。$i\in\{0,1\}$,当$i=1$时表示$\boldsymbol{x}_1$和$\boldsymbol{x}_2$是正样本对,当$i=0$时表示x_1和x_2是负样本对。边界指标$m\in\{0,1,2\}$表示$\boldsymbol{x}_1$和$\boldsymbol{x}_2$所在向量空间的可分离性:当$m=0$时,表示负样本对的损失不影响网络参数更新;当$m>0$时正负样本对的损失都影响网络参数更新。m=1和m=2表示负样本对损失函数的影响程度,m越大,负样本反向传播的权重越大。关于正负样本的选择请参考SPGAN。

总损失函数如下所示:

$$L_{\text{sp}}=L_{\text{Tadv}}+L_{\text{Sadv}}+\partial_1L_{\text{cyc}}+\partial_2L_{\text{ide}}+\partial_3L_{\text{con}} \quad (6)$$

总损失函数分为3个部分:生成器、鉴别器和孪生网络,3个部分的参数交替更新。

1.2　时空模型

现实场景中,摄像头除了记录图像信息外,还记录着图像的拍摄时间以及自身编号。图像的拍摄时间和所拍摄的摄像头信息即是图像的时间及空间信息。

TFusion算法假设两个样本间的时空点属于同一行人的概率可以表示为如下形式:

$$P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|\gamma(T_i)=\gamma(T_j)\right) \quad (7)$$

其中:T_i表示行人在t_i时刻由摄像头c_i所拍摄的样本,T_j表示行人在t_j时刻由摄像头c_j所拍摄的样本。$\gamma(T_i)=\gamma(T_j)$表示样本图像T_i和T_j属于同一行人,$\Delta_{ij}=t_i-t_j$表示两样本间的拍摄时间差。

给定任意样本对T_i和T_j,假设存在图像分类器可以判断$T_i\cong_C T_j$或者$T_i\ncong_C T_j$,即是判断两个样本是否属于同一行人。因此,对目标数据集样本进行判断,可得$P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\cong_C T_j\right)$和$P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\ncong_C T_j\right)$,式(7)可以由下式求得:

$$\begin{aligned}&P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|\gamma(T_i)=\gamma(T_j)\right)=\\&(1-E_{\text{n}}-E_{\text{p}})^{-1}\left[(1-E_{\text{n}})*P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\cong_C T_j\right)\right]-\\&E_{\text{p}}*P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\ncong_C T_j\right)\end{aligned} \quad (8)$$

其中:E_{n}和E_{p}表示图像分类器的误差率(下标p和n分别表示positive和negative),当误差率趋近于0时$P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\cong_C T_j\right)$趋近式(7)。关于式(8)的详细证明过程请参考TFusion附录部分。

1.3　贝叶斯融合方法

融合模型的计算公式如下所示:

$$P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j,\Delta_{ij},c_i,c_j\right) \quad (9)$$

其中$\boldsymbol{v}_i$和$\boldsymbol{v}_j$表示样本T_i和T_j经过图像分类器提取后的特征。式(9)表示给定样本的视觉特征及时空信息的条件下,样本T_i和T_j属于同一个人的概率。

根据贝叶斯定理,式(9)可以等价于:

$$\begin{aligned}&P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j,\Delta_{ij},c_i,c_j\right)=\\&\frac{P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right)*P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|\gamma(T_i)=\gamma(T_j)\right)}{P_{\text{r}}(\Delta_{ij},c_i,c_j)}\end{aligned} \quad (10)$$

其中$P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right)$表示已知视觉特征的条件下,样本$T_i$和$T_j$属于同一个人的概率,可由下列公式计算得到:

$$\begin{aligned}&P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right)=\\&(1-E_{\text{n}}-E_{\text{p}})*P_{\text{r}}\left(T_i\cong_C T_j\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right)+E_{\text{n}}\end{aligned} \quad (11)$$

其中$P_{\text{r}}\left(T_i\cong_C T_j\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right)$为图像分类器$C$的计算结果,$E_{\text{n}}$和$E_{\text{p}}$是图像分类器的误差率。

根据式(10)与式(11)可得:

$$\begin{aligned}&P_{\text{r}}\left(\gamma(T_i)=\gamma(T_j)\middle|\boldsymbol{v}_i,\boldsymbol{v}_j,\Delta_{ij},c_i,c_j\right)=\\&\frac{\left(Z_1+\dfrac{E_{\text{n}}}{1-E_{\text{n}}-E_{\text{p}}}\right)\left[(1-E_{\text{n}})Z_2-E_{\text{p}}Z_3\right]}{P_{\text{r}}(\Delta_{ij},c_i,c_j)}\end{aligned} \quad (12)$$

其中Z_1、Z_2和Z_3定义如下:

$$Z_1=P_{\text{r}}\left(T_i\cong_C T_j\middle|\boldsymbol{v}_i,\boldsymbol{v}_j\right) \quad (13)$$

$$Z_2=P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\cong_C T_j\right) \quad (14)$$

$$Z_3=P_{\text{r}}\left(\Delta_{ij},c_i,c_j\middle|T_i\ncong_C T_j\right) \quad (15)$$

2　基于生成对抗网络联合时空模型算法

基于生成对抗网络联合时空模型(STUGAN)算法通过引入生成对抗网络生成目标场景的训练样本,增强识别模型的稳定性;利用时空特征构建目标场景样本的时空模型,筛选低概率匹配样本,提高算法准确率。

2.1　算法思想

假定源数据集$S=\{S_1,S_2,\cdots,S_n\}$与目标数据集$T=\{T_1,T_2,\cdots,T_m\}$,不包含同一行人,其中S_i标签信息包含行人编号、摄像头编号和图像所拍摄的时间序列。T_j标签信息包含摄像头编号和时间序列。

由于目标场景的数据集不具备标签信息,无法进行有监督特征提取模型(图像分类器)的训练。因此本文使用生成对抗网络进行图像生成,在保留数据集S中行人主体信息的同时将背景风格转化为目标数据集场景的风格的数据集$G(S)$,以此减少图像背景信息对于图像分类器的影响。

假定行人在各个摄像头之间的运动存在一定的时间规律,则可通过行人的运动规律筛选错误样本对,如两张视觉特征极其相似的图像(非同一行人)被图像分类器判定为同一行人,但两张图像的时间差相差较大,则可以通过时空模型进行错样本对的矫正。由式(8)可知要得到数据集T的时空模型,需要知道同一行人的图像,但由于数据集T不包含行人的标签信息,因此本文利用图像分类器对数据集T中的图像进行识别,构建目标数据集的弱标签信息,以此构建时空模型。

算法由三部分组成,算法流程框架如图2所示。

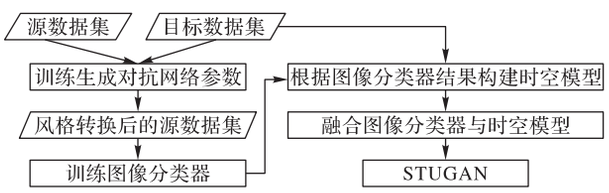


图2　算法框架

Fig. 2　Framework of algorithm

图像分类器的损失函数如下:

www.joca.cn

$$L_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log p_{ik} + \lambda\|\boldsymbol{W}\|_2 \tag{16}$$

其中：y_{ik}表示真实标签信息，p_{ik}表示模型预测的样本标签信息。正则化参数$\lambda = 0.0005$，其中$\boldsymbol{W}_i$的维度等于输出维度，$\boldsymbol{W} = \{\boldsymbol{W}_1, \boldsymbol{W}_2, \cdots, \boldsymbol{W}_{1024}\}$。

图像分类器作为中间组件，将生成对抗网络和时空模型联合。生成对抗网络提升图像分类器对于目标数据的识别准确率，使得在构建时空模型过程中图像分类器的识别误差率减小，从而构建更接近于真实情况时空模型。

2.2 算法描述

算法训练和测试过程的详细步骤如下：

训练过程：

输入：源数据集S，目标数据集T
初始化：超参数$\partial_1 = 10$，$\partial_2 = 5$，$\partial_3 = 2$，$m = 2$
For epoch=1 to 5
  根据式(6)更新生成对抗网络的网络参数
End
输出：生成数据集$G(S)$
输入：数据集$G(S)$及其标签信息
初始化：$\lambda = 0.0005$
For epoch=1 to 40
  根据式(16)更新图像分类器的网络参数
End
输出：图像分类器模型
输入：目标数据集T
For each T_i in T
  For each T_j in T
    图像分类器判断T_i与T_j的特征相似程度
  End
  相似程度由高到低排序
End
根据式(8)计算数据集T的时空模型
输出：目标数据集的时空模型

测试过程：

输入：数据集T的测试数据集query和test
For each T_i in query
  For each T_j in test
    根据式(10)计算融合识别结果
  End
End
输出：每位行人的匹配结果

根据式(8)得知，图像分类器的误差率影响时空模型的真实性，当图像分类器预测的标签信息越准确，所构建的时空模型更加接近真实情况。时空模型越接近于真实情况，能够更有效地筛选出错误的相似样本，提高总体模型在目标场景中的表现。

3 实验验证与分析

3.1 实验设置

1）网络参数。

生成对抗网络的参数设置参考SPGAN论文作者设定的参数，其中式(6)的∂_1、∂_2和∂_3分别设置为10、5和2，式(5)中的m设置为2。算法的初始学习率为0.0002，总共训练5个epoch。

图像分类器使用在ImageNet[26]预训练的ResNet-50[27]作为特征提取模型，根据训练数据的ID数量设置输出层的维度。图像分类器的训练过程中，冻结ResNet-50网络中BN层、conv1层以及res2层的参数，网络输出层的参数采用高斯分布初始化，网络参数反向传播的优化算法采用随机梯度下降法(Stochastic Gradient Descent，SGD)。超参数的设定参考文献[20]特征识别模型的训练超参数设定经验，将每一批次batch中的样本数量、最大迭代次数和动量分别设置为16、50和0.9。学习率采用阶梯型衰减策略，初始学习率为0.001，在40个epoch之后学习率衰减为0.0001，直至训练结束。

由2.2节可知，构建时空模型时需要判断样本对是否属于同一行人，本文算法直接假设图像分类器的误差率为0，即是令$E_n = E_p = 0$代入式(8)求得时空模型。

2）数据集。

为了验证算法的有效性，本文选取的两个数据集分别是Market-1501和DukeMTMC-reID，实验过程中交替使用其中一个数据集作为源数据集，剩下的数据集则作为目标数据集。关于两个数据集的具体信息如表1所示。

表1 数据集介绍

Tab. 1 Introduction of datasets

数据集	数据	样本数	ID数
Market-1501	train	12936	751
	test	19732	750
	query	3368	750
DukeMTMC-reID	train	16522	702
	test	17661	702
	query	2228	702

每个数据集都由两个部分组成，训练数据集和测试数据集，训练数据集中的行人不会出现在测试数据集。待匹配样本库(query)是从备选库(test)所属ID行人中抽取部分样本组成，两个库共同组成了测试数据集。数据集中每个样本的标签信息包含行人的ID、样本所拍摄的摄像头信息和拍摄时间三部分。Market-1501由32668个样本组成，包含6个摄像头的数据，总共1501位行人。DukeMTMC-reID由36411个样本组成，包含8个摄像头的数据，总共1401位行人。

3）评价指标。

实验分为两个部分：第一部分为图像分类器的比较，为了验证生成样本对图像分类器的影响，通过与TFusion算法的图像分类器进行比较实验，实验分析将在3.2.1节中详细阐述；第二部分为STUGAN算法与其他算法的比较，在选取的数据集中进行实验，通过与BoW(Bag-of-Words)[24]、PUL(Progressive Unsupervised Learning)[28]、UMDL(Unsupervised Multi-task Dictionary Learning)[16]、基于聚类的非对称度量学习(Clustering-based Asymmetric MEtric Learning，CAMEL)[17]、SPGAN[20]和TFusion[23]算法进行对比，以验证本文算法的性能表现。其中，算法BoW采用线性搜索策略，从数据集的表现结果中寻找最优参数，本文选取BoW算法在目标数据集的最优情况的准确率作为对比；PUL通过随机梯度下降法更新卷积神经网络(Convolutional Neural Network，CNN)识别模型的参数，然后对目标数据集样本特征进行聚类，根据聚类结果重新训练CNN模型，微调网络参数；UMDL通过拉格朗日对偶法和分析法求解多个投影矩阵的最优参数；CAMEL利用拉格朗日法将求解过程转化为因式分解问题，求解最优化投影矩阵参数；SPGAN和TFusion的优化过程是利用损失函数计算预测标签和真实标签的差值，通过随机梯度下降法更新识别

www.joca.cn

网络中的参数。

行人重识别算法采用定量衡量作为衡量标准,采用rank-1、rank-5、rank-10及平均精度均值(mean Average Precision, mAP)[24]作为行人重识别算法的表现指标。所有指标的范围为0%~100%,指标值越大,表明算法效果越好。

4)实验条件。

本文算法中的所有模型都在一台有着GTX1080Ti GPU的Linux服务器上进行训练和测试。Market-1501作为源数据集时,其中SPGAN训练时长为103 min,图像分类器训练时长为78 min,时空模型构建时长为145 min。假设测试样本数量为N,待匹配样本数量为M,测试过程中图像分类器使用的是矩阵运算计算样本间的特征相似度,时间复杂度为$O(N*M)$,通过利用GPU加速可以提高运算效率。时空模型的运算时间相当于字典的检索,时间复杂度为$O(N*M)$,算法总时间复杂度为$O(N*M)$。

3.2　算法对比分析

3.2.1　图像分类器对比分析

表2为图像分类器在Market-1501和DukeMTMC-reID数据集上的性能表现情况,其中TFusion'C和STUGAN'C分别表示TFusion算法和本文算法的图像分类器。当测试集为Market-1501时,训练数据集为DukeMTMC-reID,反之亦然。TFusion'C(SPGAN)和STUGAN'C(SPGAN)表示使用生成数据训练后的图像分类器,生成的数据来自DukeMTMC-reID的训练数据集。

表2　图像分类器在不同数据集上的性能表现　　单位:%

Tab. 2　Visual classifier performance on different datasets　unit:%

数据集	图像分类器	rank-1	rank-5	rank-10	mAP
Market-1501	TFusion'C	37.6	56.8	64.5	16.0
	STUGAN'C	40.3	59.3	67.7	16.1
	TFusion'C(SPGAN)	36.7	55.9	64.7	15.8
	STUGAN'C(SPGAN)	**46.9**	**65.6**	**73.0**	**20.1**
DukeMTMC-reID	TFusion'C	28.4	43.5	52.0	14.7
	STUGAN'C	33.4	48.7	55.6	17.4
	TFusion'C(SPGAN)	29.9	48.3	54.8	15.4
	STUGAN'C(SPGAN)	**36.1**	**51.3**	**58.1**	**18.2**

从表2中得出,本文的图像分类器无论在直接迁移或使用生成数据集的条件下,在两个数据集的表现均优于TFusion算法的图像分类器。本文图像分类器效果更好的原因在于训练过程中冻结了低层卷积的参数,使得高层的行人全局特征提取层得到更好的训练。因为生成样本的图像视觉风格更接近于目标场景的数据风格,一定程度上减少了场景转换对图像分类器的影响。

3.2.2　总体模型对比分析

表3为各算法基于Market-1501和DukeMTMC-reID数据集的rank-1、rank-5、rank-10及mAP指标表现情况。其中算法SPGAN、TFusion、STUGAN存在训练过程,当测试集为Market-1501时,训练数据集为DukeMTMC-reID,反之相同。

从表3中可得出在两个数据集的rank-1、rank-5及rank-10指标比较中,本文算法的准确率均优于所比较的算法。尤其在数据集Market-1501上,rank-1指标的准确率相比TFusion算法高出5.7个百分点。其中BoW为手工设计的算法,因为没有任何训练过程,因此算法在两个数据集的准确率都偏低。PUL和UMDL为无监督算法,因为缺少标签信息进行算法的学习,所以算法的准确率较低。与使用了生成对抗网络的SPGAN算法相比,本文算法在此基础上加入了时空模型的辅助,时空模型的加入一定程度上排除视觉特征高度相似的错样本被匹配,减小了图像分类器的误差率。对于使用了时空模型作为辅助的TFusion算法,本文算法的图像分类器更好地利用生成对抗网络生成的样本进行训练,减少场景之间领域偏差对图像分类器的负面影响。较高水平的图像分类器构建的时空模型更接近于真实情况,两者的联合使得STUGAN算法在目标场景中具备良好的识别性能。

表3　不同算法在Market-1501和DukeMTMC-reID数据集上的性能对比结果　　单位:%

Tab. 3　Algorithm performance comparison results on Market-1501 and DukeMTMC-reID datasets　unit:%

数据集	算法	rank-1	rank-5	rank-10	mAP
Market-1501	BoW	35.8	52.4	60.3	14.8
	PUL	45.5	60.7	66.7	20.5
	CAMEL	54.5	—	—	26.3
	UMDL	34.5	52.6	59.6	20.5
	SPGAN	57.0	73.9	80.3	**27.1**
	TFusion	60.7	73.7	79.3	18.0
	STUGAN	**66.4**	**78.9**	**84.7**	21.5
DukeMTMC-reID	BoW	17.1	28.8	34.9	8.3
	PUL	18.5	31.4	37.6	7.3
	UMDL	30.0	43.4	48.7	16.4
	SPGAN	41.1	56.6	63.0	**22.3**
	TFusion	50.9	63.7	67.3	19.5
	STUGAN	**53.5**	**69.0**	**74.0**	20.0

4　结语

本文针对行人重识别算法跨场景应用时出现算法准确率大幅下降问题,提出一种基于生成对抗网络联合时空模型的识别方法。算法在目标场景中获得相对满意的表现,但是距离实际应用要求的准确率仍存在一定的距离。由于图像分类器的识别准确率相对较低,在构建时空模型时存在较多的错误样本对被判别为同一行人,导致时空模型与真实的时空模型具有一定的偏差。如何构建适应于场景变化的图像分类器,使其在跨场景应用中仍然保持良好的准确率是今后的研究方向。

参考文献(References)

[1] 宋婉茹,赵晴晴,陈昌红,等. 行人重识别研究综述[J]. 智能系统学报, 2017, 12(6):770-780. (SONG W R, ZHAO Q Q, CHEN C H, et al. Survey on pedestrian re-identification research [J]. CAAI Transactions on Intelligent Systems, 2017, 12(6):770-780.)

[2] OJALA T, PIETIKÄINEN M, MÄENPÄÄ T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987.

[3] SWAIN M J, BALLARD D H. Color indexing [J]. International Journal of Computer Vision, 1991, 7(1): 11-32.

[4] GRAY D, TAO H. Viewpoint invariant pedestrian recognition with an ensemble of localized features [C]// Proceedings of the 2008 European Conference on Computer Vision, LNCS 5302. Berlin: Springer, 2008: 262-275.

[5] KVIATKOVSKY I, ADAM A, RIVLIN E. Color invariants for person reidentification [J]. IEEE Transactions on Pattern Analysis

www.joca.cn

and Machine Intelligence, 2013, 35(7): 1622-1634.

[6] LI W, WANG X. Locally aligned feature transforms across views [C]// Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2013: 3594-3601.

[7] LIAO S, HU Y, ZHU X, et al. Person re-identification by local maximal occurrence representation and metric learning [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 2197-2206.

[8] AHMED E, JONES M, MARKS T K. An improved deep learning architecture for person re-identification [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 3908-3916.

[9] CHEN S, GUO C, LAI J. Deep ranking for person re-identification via joint representation learning [J]. IEEE Transactions on Image Processing, 2016, 25(5): 2353-2367.

[10] 陈兵,查宇飞,李运强,等. 基于卷积神经网络判别特征学习的行人重识别[J]. 光学学报, 2018, 38(7): No. 0720001. (CHEN B, ZHA Y F, LI Y Q, et al. Person re-identification based on convolutional neural network discriminative feature learning [J]. Acta Optica Sinica, 2018, 38(7): No. 0720001.)

[11] KÖSTINGER M, HIRZER M, WOHLHART P, et al. Large scale metric learning from equivalence constraints [C]// Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2012: 2288-2295.

[12] TAO D, GUO Y, SONG M, et al. Person re-identification by dual-regularized kiss metric learning [J]. IEEE Transactions on Image Processing, 2016, 25(6): 2726-2738.

[13] PAISITKRIANGKRAI S, SHEN C, VAN DEN HENGEL A. Learning to rank in person re-identification with metric ensembles [C]// Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2015: 1846-1855.

[14] YI D, LEI Z, LIAO S, et al. Deep metric learning for person re-identification [C]// Proceedings of the 22nd International Conference on Pattern Recognition. Piscataway: IEEE, 2014: 34-39.

[15] CHEN Y, ZHENG W, LAI J, et al. An asymmetric distance model for cross-view feature mapping in person reidentification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 27(8): 1661-1675.

[16] PENG P, XIANG T, WANG Y, et al. Unsupervised cross-dataset transfer learning for person re-identification [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 1306-1315.

[17] YU H, WU A, ZHENG W. Cross-view asymmetric metric learning for unsupervised person re-identification [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 994-1002.

[18] GENG M, WANG Y, XIANG T, et al. Deep transfer learning for person re-identification[EB/OL]. [2019-10-18]. https://arxiv.org/pdf/1611.05244.pdf.

[19] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// Proceedings of the 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2017: 2242-2251.

[20] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 994-1003.

[21] HUANG W, HU R, LIANG C, et al. Camera network based person re-identification by leveraging spatial-temporal constraint and multiple cameras relations [C]// Proceedings of the 2016 International Conference on Multimedia Modeling, LNCS 9516. Cham: Springer, 2016: 174-186.

[22] MARTINEL N, FORESTI G L, MICHELONI C. Person reidentification in a distributed camera network framework [J]. IEEE Transactions on Cybernetics, 2016, 47(11): 3530-3541.

[23] LV J, CHEN W, LI Q, et al. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns [C]// Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 7948-7956.

[24] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark [C]// Proceedings of the 2015 IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1116-1124.

[25] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking [C]// Proceedings of the 2016 European Conference on Computer Vision, LNCS 9914. Cham: Springer, 2016: 17-35.

[26] DENG J, DONG W, SOCHER R, et al. ImageNet: a large-scale hierarchical image database [C]// Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2009: 248-255.

[27] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[28] FAN H, ZHENG L, YAN C, et al. Unsupervised person re-identification: Clustering and fine-tuning [J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(4): No. 83.

This work is partially supported by the National Key Research and Development Program of China (2018YFB1802400), the Guangdong Research and Development Planning Project of Key Fields (2019B010118001, 2019B010121001).

QIU Yaoru, born in 1993, M. S. candidate. His research interests include deep learning, image recognition.

SUN Weijun, born in 1975, Ph. D., lecturer. His research interests include intelligent information processing.

HUANG Yonghui, born in 1975, Ph. D., lecturer. Her research interests include image processing, pattern recognition.

TANG Yuqi, born in 1995, M. S. candidate. Her research interests include deep learning, image recognition.

ZHANG Haochuan, born in 1994, M. S. candidate. His research interests include deep learning, computer vision.

WU Junpeng, born in 1994, M. S. candidate. His research interests include machine learning, pattern recognition.

www.joca.cn