第 59 卷 第 10 期/2022 年 5 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

基于生成对抗网络与自校准卷积的行人重识别

李开放[1]，惠冠程[1]，王汝涵[1]，张苗辉[1,2*]

[1]河南大学人工智能学院，河南 开封 475004；

[2]河南大学大数据分析与处理河南省重点实验室，河南 开封 475004

摘要 针对行人重识别过程中跨相机拍摄导致的行人图像风格差异问题，提出了一种基于循环矢量量化生成对抗网络(CVQGAN)与自校准卷积模块的学习框架。设计了一种离散化的矢量量化模块，将该模块用于生成器由编码到解码的过程中，利用矢量量化空间中的离散矢量解决了原始生成器产生噪声伪图像的问题，从而生成质量更高的风格转换图像。将自校准卷积模块融合至Resnet50主干网络的卷积层中，利用多分支网络结构对各支路进行不同的卷积操作，以获取表征能力更强的特征，进一步解决同一行人在不同相机下的风格差异问题。在Market1501和DukeMTMC-reID数据集上对所提算法进行有效性实验验证，结果表明本文算法能够有效提高行人重识别的准确率和鲁棒性。

关键词 机器视觉；跨相机；生成对抗网络；风格转换；自校准卷积；行人重识别

中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP202259.1015007

Person Re-Identification Based on Generative Adversarial Network and Self-Calibrated Convolution

Li Kaifang[1], Hui Guancheng[1], Wang Ruhan[1], Zhang Miaohui[1,2*]

[1]*School of Artificial Intelligence, Henan University, Kaifeng* 475004, *Henan, China*;

[2]*Henan Key Laboratory of Big Data Analysis and Processing, Henan University, Kaifeng* 475004, *Henan, China*

Abstract Aiming at the problem of person image style difference caused by cross-camera shooting in the process of person re-identification, this paper proposes a learning framework based on a cyclic vector quantization generative adversarial network (CVQGAN) and a self-calibrated convolution module. First of all, this paper designs a discrete vector quantization module, which is introduced into the process from encoding to decoding of the generator. The discrete vector in the vector quantization space is used to solve the problem that the original generator produces noisy pseudo images, therefore generating higher quality style conversion images. Then, the self-calibration convolution module is integrated into the convolution layer of the Resnet50 backbone network, and the multi-branch network structure is used to perform different convolution operations on each branch, so as to obtain features with stronger characterization ability and further solve the problem of style differences of the same pedestrian under different cameras. The proposed algorithm is validated by experiments on Market1501 and DukeMTMC-reID datasets, and the results show that the proposed algorithm can effectively improve the accuracy and robustness of person re-identification.

Key words machine vision; cross camera; generative adversarial networks; style transfer; self-calibrated convolution; person re-identification

收稿日期：2021-04-20；**修回日期**：2021-05-13；**录用日期**：2021-05-25

基金项目：国家自然科学基金(61802111,62002100)、河南省教育厅科学技术研究重点项目(19A50002)

通信作者：*zhmh@henu.edu.cn

1 引 言

行人重识别[1]，也称行人再识别，是指给定一个摄像头拍摄的行人图像，从其他视野可能重叠但视角不同的多个摄像头捕获的大量图像中重新识别该行人的过程，也可将其理解为图像检索，其在智能视频监控、刑侦等领域有着非常广阔的应用前景，也是近年来计算机视觉领域的研究热点。

在实际监控环境中，行人重识别面临着遮挡、跨模态和样本量少等问题。不同位置部署的摄像头具有较大的环境差异，拍摄到的行人图片往往背景杂乱且存在遮挡问题。跨模态的行人重识别主要为了解决行人的RGB图像和红外图像等不同模态下图像的交叉模态变化问题。相对于有监督学习，无监督的行人重识别任务不需要大量的有标签样本，其主要挑战在于学习样本图像中无标签的判别性特征识别。在实际场景中，现有技术无法有效地解决上述的各种挑战，行人重识别任务依然是国内外专家学者高度关注和广泛研究的重点。

为应对背景、光照和分辨率等因素造成相机采集到的行人图像在外观和风格上有差异的问题，文献[2-5]尝试用不同的方法去解决行人图像风格差异的问题。经典的方法包括KissMe[2]和XQDA[3]等，其中，KissMe算法使用似然比检验来判断两张图片之间的差异程度；XQDA算法利用高斯模型分别拟合类内和类间样本特征的差值分布，再使用对数似然比推导出马氏距离。深度学习的方法包括SVDNet[4]和TripletNet[5]等，其中，SVDNet算法利用正交性约束提升特征向量的表达能力；TripletNet算法由三个相同且彼此参数共享的前馈神经网络组成，分别计算正样本、负样本与候选样本的欧氏距离。以上方法都是在不同相机之间提取同一行人的不变性特征，但往往无法充分挖掘样本分布中更加丰富的其他特征信息。

另外一种思路则是通过扩充数据集的方式减小图像风格的差异性。但这种方法仍然存在一个问题，虽然利用扩充后的数据集能够提升识别的各项指标，但大规模的人工标注成本非常高。为了解决人工标注问题，文献[6-7]提出了多种数据扩充和正则化方法。其中，文献[6]使用DCGAN[8]生成未标记的样本，并为它们分配统一的标签以提高CNN模型的辨别能力。与文献[6]相反，Zhu等在文献[8]中提出的CycleGAN实现了对不同风格的图像进行转换，且风格转换的样本是从真实数据中产生的。因此只向训练集中添加更多的样本，而不重新标注新的数据，这样既能解决数据少的问题，也避免了标注成本的增加。除此之外，与此类似的生成对抗网络还有DualGAN[9]和DiscoGAN[10]等。

同时，行人重识别的准确率很大程度上也取决于行人的特征信息，行人特征信息获取越全面，重识别的效果就越好。文献[11]提出了一种多尺度卷积特征融合算法，使用金字塔池化方法获得全局特征和多尺度局部特征，以提升特征的鉴别能力。Simonyan等[12]提出的VGGNet使用更小核尺寸(3×3)的卷积滤波器来构建更深层次的网络，从而使得网络在使用更少参数的情况下具有更好的性能；毕晓君等[13]提出了一种基于视角信息嵌入的行人重识别模型，利用行人图像视角朝向特点对视角单元进行特征提取，以进一步优化网络；Chen等[14]提出了一种级联抑制策略，使网络更多地挖掘被显著特征掩盖的各种潜在的有用特征。

针对行人重识别过程中跨相机拍摄导致的行人图像风格差异问题以及传统卷积结构进行卷积操作时感受野较小导致的鲁棒性和判别力较差的问题，本文提出了一种基于生成对抗网络与自校准卷积的行人重识别学习框架。

2 基本原理及网络结构

2.1 基本原理

GAN(generative adversarial network)最初是由Goodfellow等[15]在2014年所提出，主要用于图像之间的翻译与转换。基本的GAN模型包含生成器G(generator)和判别器D(discriminator)两个网络，网络示意图如图1所示，整个GAN的目标函数可表示为

$$\min_G \max_D V(D,G)=E_{x\sim P_{data}(x)}[\lg D(x)]+E_{z\sim P_z(x)}\lg\{1-D[G(z)]\},\tag{1}$$

式中：x为真实样本数据；数据分布为P_{data}；生成器的输入为P_z的随机噪声$\boldsymbol{Z}$，然后输出$\boldsymbol{Z}$到真实数据空间的映射$G(\boldsymbol{Z};\theta_g)$，$\theta_g$为生成器参数；判别器网络为$D(\boldsymbol{Z};\theta_d)$，$\theta_d$为判别器参数；$D(x)$为真实输入的概率；$G(z)$表示生成器生成的图像。在训练过程中，两个网络交替进行训练。训练生成器时，固定判别器参数，极小化$V(D,G)$；训练判别器时，固定生成器参数，极大化$V(D,G)$。两者在不断博弈中对网络进行优化，使最终产生的图片越来越接近真实图片。经

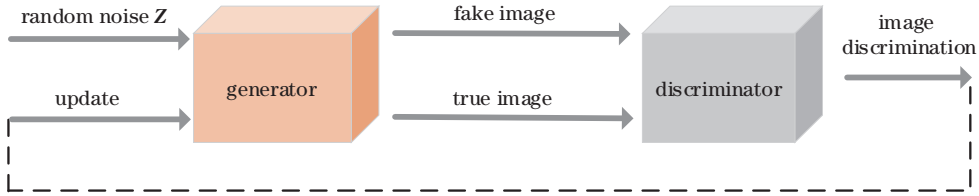

图 1　GAN 示意图

Fig. 1　GAN diagram

过近些年的发展，GAN 已经取得了很大的成功，特别是在图像生成方面有很多改进形式，如 DCGAN[7]，CycleGAN[8]，DualGAN[9]，DiscoGAN[10]等。

CycleGAN 的本质是两个镜像对称的 GAN 构成的一个环形网络，两个镜像 GAN 一共拥有两个生成器和两个判别器[8]，如图 2(a)所示。采用两个生成器是为了避免所有的数据域 X 都被映射到同一个数据域 Y 上，这样既能满足 $X\rightarrow Y$ 的映射，也能满足 $Y\rightarrow X$ 的映射。同时，采用两个对抗性判别器 D_Y 和 D_X 对图像生成过程进行监督，以实现模型的最优参数。为了进一步正规化两个生成器相互间的映射，引入了两个循环一致性损失，即图 2(b)中的前向循环一致性损失 $X\rightarrow G(X)\rightarrow F\left[G(X)\right]\approx\hat{x}$，以及图 2(c)中的后向循环一致性损失 $Y\rightarrow F(Y)\rightarrow G\left[F(Y)\right]\approx\hat{y}$，其中，$\hat{X}$ 与 $\hat{Y}$ 分别表示由真实域 Y 与 X 经模型映射生成的目标域数据，$\hat{x}$ 与 $\hat{y}$ 则是目标域数据再次经模的逆映射生成的近似源域数据。

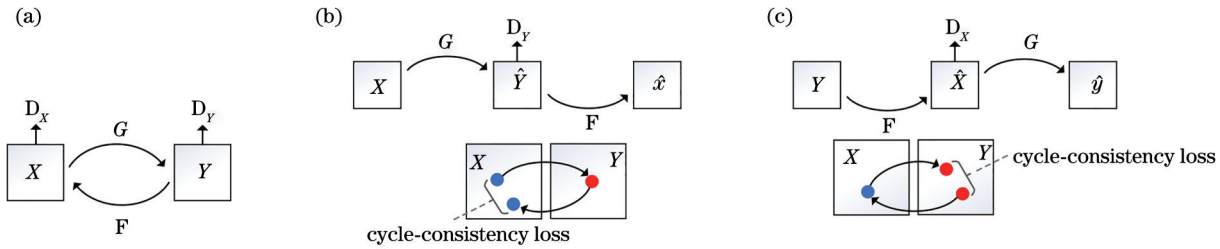

图 2　环形网络工作原理。(a)生成器及判别器；(b)前向循环一致性损失；(c)后向循环一致性损失

Fig. 2　Cyclic network working principle. (a) Generator and discriminator; (b) forward cycle-consistency loss; (c) backward cycle-consistency loss

文献[16]正是基于这种思想提出了 CamStyle 方法，但是仍然存在一些问题：1)CycleGAN 生成的 Camstyle 图像样本中会有图像噪声伪影，导致产生错误图像，如图 3 所示，Cam i (i=1,2,3,4,5,6)分别为 Market-1501 数据集下 6 个不同视角的摄像头，其中标志处为转换过程中产生的图像噪声；2)生成器中由编码到解码的转换过程采用 Resnet 残差模块，卷积层数量较多，且需要训练的模型数量为 C_C^2 个，其中 C 为数据集中摄像机的数量，因此该过程不适用于计算资源不足的场景。本文提出的循环矢量量化生成对抗网络(CVQGAN)解决了 CycleGAN 产生伪图像的问题，生成了质量更高且风格统一的 CVQStyle 图像，并且训练所需要的计算资源也更少。

类似于 GAN，Variational autoencoder(VAE)是 2013 年由 Kingma 等[17]提出的一种基于变分思想的深度学习生成模型。它的目标与 GAN 的目标基本相同，都是希望构建一个从源数据 X 生成目标数据 Y 的模型。VAE 又称为变分自编码器，由两个部分组成，即 encoder 编码器网络和 decoder 解码器网络，可以将源域的原始数据转换为不同风格的目标域数据。随机输入一个给定数据分布的 n 维向量，用于生成一张新的图片。对于 GAN，生成器通过学习真实样本的数据分布进行训练以生成伪造样本，而判别器则对真实与伪造样本进行概率估计，两者通过对抗学习的方式获得较好的模型效果。不同于 GAN 生成图像的方式，对于 VAE 而言，n 维向量代表的是 n 个决定最终生成图片样式的隐形因素。每一个因素都对应着一种分布，先从这些分布中进行采样，再通过深度网络恢复图片。

2.2　CVQGAN 模型

本文利用 VQ(vector quantization)的思想，设计

图 3　由 Market-1501 中的 CycleGAN 和 CVQGAN 生成的示例

Fig. 3　Examples generated by CycleGAN and CVQGAN in Market-1501

了一种离散化的矢量量化模块，将该模块融入 CycleGAN 中，取代原始生成器结构中的 Resnet 转换模块。CVQGAN 生成器结构流程图如图 4 所示，首先定义一个潜在的矢量量化空间 $\boldsymbol{v}\in\mathbf{R}^{K\times N}$，$K$ 为离散潜在空间的大小，N 为每个潜在离散向量 $\boldsymbol{v}_i$ 的维数，即存在 K 个离散向量 $\boldsymbol{v}_i\in\mathbf{R}^N$，$i=1,2,\cdots,K$，因此其 code 不再是由 encoder 直接输出得到的连续码，而是经过一个矢量量化后得到的离散码，这对解决一些实际问题更加有帮助。

首先输入 x，数据结构为 $[B,3,H,W]$，其中 B 为 batch 的数量，编码器输入的 Channel 数为 3，H、W 则表示输入图像的长和宽。图片 x 经过 encoder 之后，会得到关于编码器深度神经网络的输出，其结构为 $[B,C=N,\hat{H},\hat{W}]$，其中 C 是指编码器的 Conv 网络

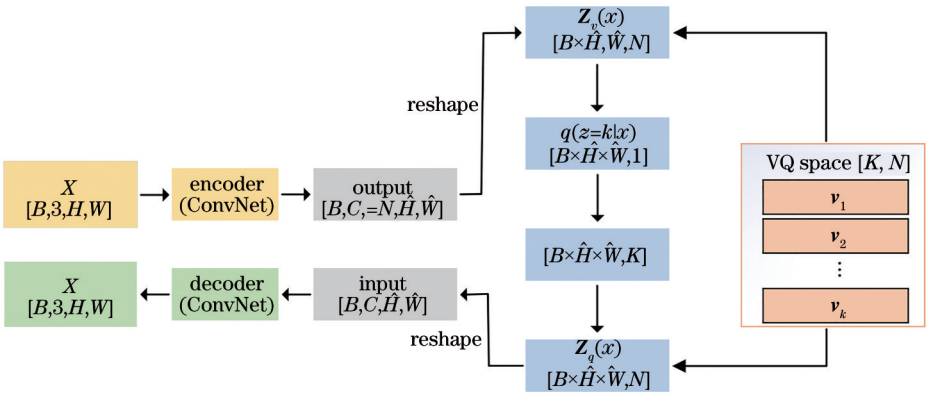

图 4　CVQGAN 生成器结构流程图

Fig. 4　Flow chart of CVQGAN generator

输出的Channel的数量，而N是指矢量量化中矢量的维度，也就是VQ space中所存储矢量的维度，$\hat{H}$、$\hat{W}$表示输入图像经编码器处理后的长和宽，再通过矢量量化空间并使用式(1)计算离散潜在随机变量z以及z的后验分布$q(z|x)$。编码器的输出经reshape后为$\boldsymbol{Z}_v(x)$，其结构为$[B\times\hat{H}\times\hat{W},N]$，即每一个图片有$\hat{H}\times\hat{W}$个编码，每个编码是$N$维，计算这些编码$\left(B\times\hat{H}\times\hat{W}\right)$与VQ space中$K$(表示矢量量化编码的矢量个数)个矢量之间的距离，通过最近邻算法构成如下映射：

$$q(z=k|x)=\begin{cases}1, k=\mathrm{argmin}_j\|\boldsymbol{Z}_v(x)-\boldsymbol{v}_j\|_2\\0, \text{otherwise}\end{cases},\quad(2)$$

式(2)表示当输入为x时，$z=k$的概率是：当k是矢量序列$\{\boldsymbol{v}_1,\boldsymbol{v}_2,\ldots,\boldsymbol{v}_k\}$中与$\boldsymbol{Z}_v(x)$最近的矢量的下标时，条件概率为1，否则为0。这里的矢量距离度量采用常见的欧拉距离$\|\cdot\|_2$，式(2)便是最近邻算法的实现。

$$\boldsymbol{z}_q(x)=\boldsymbol{v}_k \text{ if } k=\mathrm{argmin}_j\|\boldsymbol{Z}_v(x)-\boldsymbol{v}_j\|_2,\quad(3)$$

式(3)表示，通过最近邻计算出与$\boldsymbol{Z}_v(x)$最近矢量的下标为k，然后通过查表将$\boldsymbol{v}_k$输出作为编码输出$\boldsymbol{Z}_q(x)$。$\boldsymbol{Z}_q(x)$作为decoder的输入，由decoder进行图像的重建。

CVQStyle图像生成网络的结构图如图5所示，首先输入图片X，经过生成器G的编码器编码并reshape后输出$\boldsymbol{Z}_v(x)$，通过共享矢量量化空间并使用最近邻算法计算离散潜在随机变量z以及z的后验分布$q(z|x)$，再利用对应的向量$\boldsymbol{v}_k$计算出$\boldsymbol{Z}_q(x)$作为解码器的输入。利用神经网络进行调参后，生成图像Y。可以看出，原始图像X与生成图像Y的风格发生了变化。生成的Y作为输入，经生成器F重构回原始输入的图像。在这个过程中，判别器D_X和D_Y起到判别作用，确保图像的风格转换。

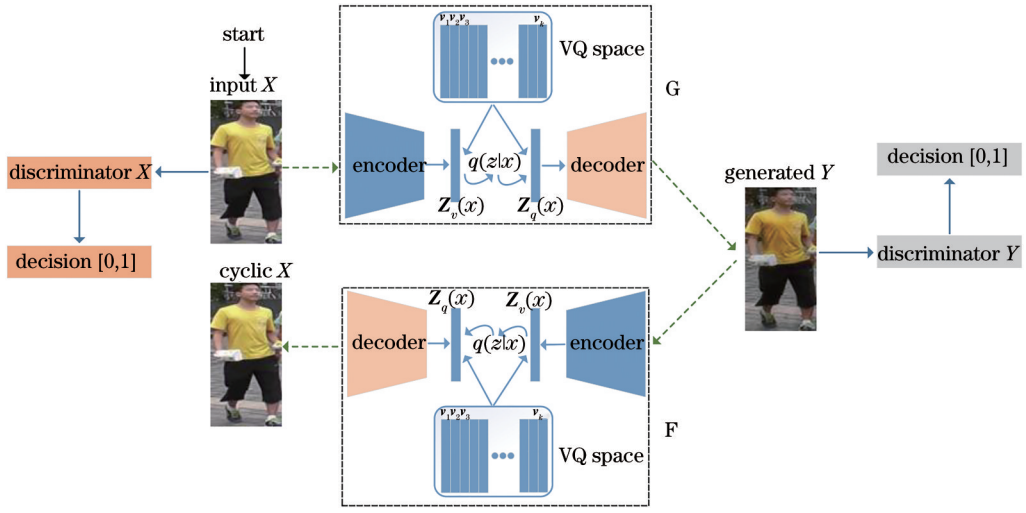

图5　CVQStyle图像生成网络

Fig. 5　CVQStyle image generation network

2.3　网络结构

本文采用残差网络结构Resnet50[18]作为本文框架的主干网络。针对传统的深度学习网络在信息传递时存在信息丢失、梯度消失或者梯度爆炸的问题，在Resnet50网络中加入残差学习的思想，通过引入一条残差边实现了跨层连接。在Resnet50网络中输入的信息可以通过残差边到达输出，这简化了神经网络训练学习的难度，也保证了信息在传输过程中的完整性，解决了梯度消失导致的深度网络退化问题。

为了进一步解决同一行人在不同相机下的风格差异问题，本文引入了自校准卷积模块SCNet[19]。自校准卷积模块网络结构图如图6所示，$\boldsymbol{X}_i$、$\boldsymbol{Y}_i(i=1,2)$分别为输入特征图与输出特征图，ξ为不同kernel size的卷积层，kernel的大小包括K_1,K_2,K_3,K_4，维度均为$(C/2)\times(C/2)\times H\times W$。

SCNet的操作主要分为两条路径：第一条路径即虚线部分为自校准操作，第二条路径为传统的卷积操作。首先，将大小为$C\times H\times W$的特征图$\boldsymbol{X}$分为$\boldsymbol{X}_1$、$\boldsymbol{X}_2$两部分。对$\boldsymbol{X}_1$进行虚线部分的自校准操作，该操作主要分为三个分支。分支1不进行操作。分支2可表示为

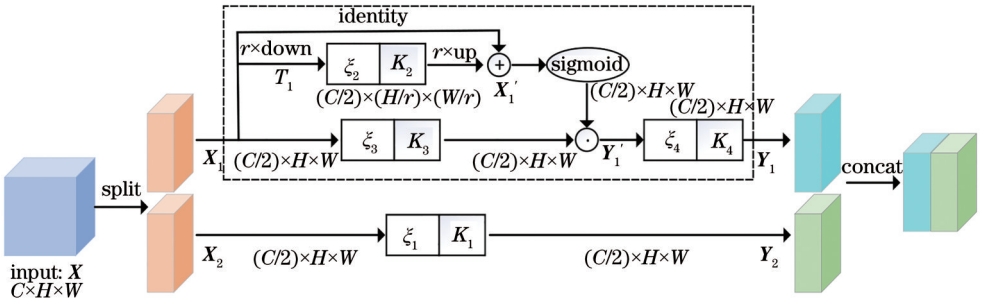

图 6　自校准卷积模块网络结构

Fig. 6　Self-calibrated convolution module network structure

$$\boldsymbol{T}_1=\mathrm{AvgPool}_r(\boldsymbol{X}_1),\tag{4}$$

式中：$\mathrm{AvgPool}_r$为平均池化操作。首先将$\boldsymbol{X}_1$进行size为$r\times r$、步长为r的平均池化下采样，得到$\boldsymbol{T}_1$、$\boldsymbol{T}_1$的大小为$(C/2)\times(H/r)\times(W/r)$；对$\boldsymbol{T}_1$进行$\xi_2$卷积后再上采样$r$倍，得到$\boldsymbol{X}_1'$：

$$\boldsymbol{X}_1'=\mathrm{Up}\left[\xi_2(\boldsymbol{T}_1)\right]=\mathrm{Up}(\boldsymbol{T}_1*\boldsymbol{K}_2),\tag{5}$$

式中：Up为双线性插值算子，用于实现上采样操作。将$\boldsymbol{X}_1$与$\boldsymbol{X}_1'$求和，经过sigmoid操作得到权重值。将权重值与分支3的结果（$\boldsymbol{X}_1$经过ξ_3卷积）相乘得到$\boldsymbol{Y}_1'$：

$$\boldsymbol{Y}_1'=\xi_3(\boldsymbol{X}_1)\cdot\sigma(\boldsymbol{X}_1+\boldsymbol{X}_1'),\tag{6}$$

式中：σ为sigmoid激活函数。

最后对$\boldsymbol{Y}_1'$经过ξ_4卷积，得到$\boldsymbol{Y}_1$：

$$\boldsymbol{Y}_1=\xi_4(\boldsymbol{Y}_1')=\boldsymbol{Y}'*\boldsymbol{K}_4。\tag{7}$$

第二条路径的目的是保留空间上下文关系，$\boldsymbol{X}_2$经过传统的卷积操作得到$\boldsymbol{Y}_2$。最后对$\boldsymbol{Y}_1$和$\boldsymbol{Y}_2$进行concat操作，得到最后的输出特征图$\boldsymbol{Y}$。对于传统的卷积模块，每个空间位置的感受野主要由预定义的卷积核大小控制。而自校准卷积模块采用多种不同size的卷积核，且考虑了空间上下关系，拥有了更大的感受野。

框架总体结构图如图7所示，对于每一张真实图片，利用训练好的CVQGAN模型来生成符合目标摄像机风格的图像。随后，将真实图像（实线框）和风格转换图像（虚线框）相结合，以训练基准模型Resnet50。本文将模型最后1000维的分类层舍弃，重新在卷积层后添加了一个输出为1024维的全连接层和一个输出为C维的全连接层，其中C为数据集中行人的ID数（如对于Market1501数据集，C=751），并且将交叉熵损失（L_{Cross}）[20]和标签平滑正则化损失（LSR，其值用L_{LSR}表示）[16]应用于真实图像和风格转换CVQStyle图像。其中，LSR损失的设计是为了应对建模过程中CVQGAN没有进行完美建模或遮挡和检测误差使真实数据存在噪声样本导致图像生成过程中出现噪声的问题。

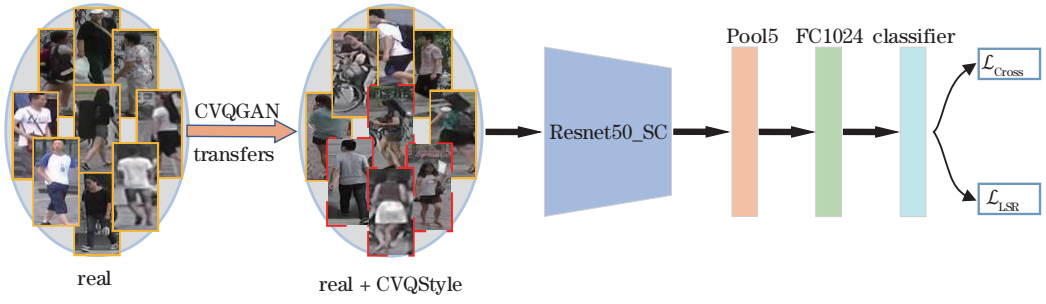

图 7　总体网络框架

Fig. 7　Overall network framework

3　实验分析与讨论

3.1　数据集及评价指标

为验证本文所提方法的有效性，在Market-1501[21]和DukeMTMC-reID[22]两个标准数据集上进行了测试实验。Market-1501数据集是在清华大学校园内采集得到的，由6个摄像头拍摄到的1501个行人组成，共32668张行人图像，每个行人的图片由2～6个相机拍摄。其中：训练集有751人，包含12936张图像；测试集有750人，包含19732张图像。

行人检测框使用DPM检测器进行标注。DukeMTMC-reID数据集是在杜克大学校园内采集，由8个不同摄像头拍摄到的1812个行人组成，但在两个以上摄像头中出现过的行人只有1404个，共34183张图像；训练集包含702个行人，共16522张图像；测试集包含702个行人，共17661张图像。行人检测矩形框由人工进行标注。这两个数据集在光照与姿态等方面均有着较大的变化，更加符合真实场景的应用。

此次实验采用首位命中率（Rank-1）与均值平均精度（mAP）两种评价指标评估算法的性能。采用了三种评价生成图像质量的指标即PSNR（peak signal to noise ratio）、FID（Frechet inception distance）和SSIM（structural similarity），对本文提出的CVQGAN与CycleGAN生成图片的质量进行比较。

3.2　实验环境及参数设置

实验使用Pytorch的深度学习框架进行网络的搭建，操作系统为Ubuntu16.04版本，编程环境为Pycharm，配备了2.50 GHz E5-2678 v3 CPU和显卡为16G的Tesla T4 GPU的设备进行网络的训练，且本文采用在ImageNet数据集上预训练的Resnet50网络作为特征提取网络。在CVQGAN的训练过程中，将所有输入图片的大小调整为256×256，并使用Adam[23]优化器对实验模型进行优化。在前30个epochs中，生成器的学习率为0.0002，鉴别器的学习率为0.0001。在剩余20个epochs中，学习率线性地降为零。对于每幅训练图像，网络将生成C_{-1}（即Market1501：5张；DukeMTMC-reID：7张）张额外的转换图像，并将其原始标签保留，作为扩充的训练数据。在训练re-ID基础模型时，将所有的输入图像尺寸调整为256×128大小，并在训练期间使用随机裁剪、随机水平翻转和随机擦除来处理训练图像。本文设置随机擦除率γ为0.5，批处理大小为64。学习率从0.01开始，并将40个epochs后的学习速率除以10，共进行了50个epochs的训练。

3.3　CVQStyle模型

本文提出的CVQGAN解决了CycleGAN产生噪声伪图像的问题，有着更出色的相机风格转换能力。相对于DCGAN的单生成器结构，CVQGAN的两个生成器G、F对数据域X和Y实现了$X\rightarrow Y$与$Y\rightarrow X$的双向映射，确保生成的图像仍然包含行人的主要特征。三种GAN的生成图像效果如图8所示。

图8　DCGAN、CycleGAN和CVQGAN生成图像示例。

Fig. 8　Image examples generated by DCGAN, CycleGAN, and CVQGAN.

为了更加直观地对比两种方法生成图像的质量，本文采用了常用的GAN生成图像质量的三种评估指标PSNR、SSIM以及FID。PSNR，又称峰值信噪比，可以更好地反映GAN生成图像过程中

产生的失真情况，其值越大则真实度更高；SSIM从亮度、对比度与结构三个方面度量两幅图像之间的相似性，以此判断生成结果的多样性，其值越大则代表模型性能越好；FID用来计算真实图像与生成图像的特征向量间距离的一种度量，其值越小，特征越相近。对图8中分别由CycleGAN与CVQGAN生成的8张风格转换图像进行比对，结果如表1所示。

由表1可以看出，本文提出的CVQGAN生成的样本图像在PSNR、SSIM和FID三项指标中均表现更好，其中image 5的PSNR和FID与image 8的SSIM均提升较高。PSNR和FID的对比数据说明了CVQStyle图像比CamStyle图像更加真实，且质量更好；SSIM更高则代表CVQStyle图像在多样性方面具有更大的竞争力。

表1　生成图像质量对比

Table 1　Generated image quality comparison

Image	Model	PSNR	SSIM	FID
Image 1	CamStyle	18.46	0.66	231.48
	CVQStyle	23.31	0.87	196.44
Image 2	CamStyle	22.52	0.79	145.70
	CVQStyle	23.71	0.91	82.80
Image 3	CamStyle	21.04	0.77	122.76
	CVQStyle	26.24	0.95	93.77
Image 4	CamStyle	17.86	0.71	223.48
	CVQStyle	22.41	0.87	180.31
Image 5	CamStyle	20.54	0.79	161.32
	CVQStyle	29.71	0.97	41.53
Image 6	CamStyle	15.07	0.52	321.54
	CVQStyle	20.11	0.82	108.99
Image 7	CamStyle	16.09	0.63	290.03
	CVQStyle	21.04	0.89	208.25
Image 8	CamStyle	13.27	0.46	282.65
	CVQStyle	20.18	0.80	132.74

3.4　消融实验

为了进一步验证本文所提方法的有效性，在Market-1501和DukeMTMC-reID两个数据集上进行了消融实验，结果如表2所示。其中，Baseline为以Resnet50为主干的基准网络，SCNet为自校准卷积模块。实验结果如表2所示，可以看到：相比实验1，实验2加入由CVQGAN生成的数据集后，Market-1501数据集的Rank-1与mAP分别提高了2.44%和3.42%，DukeMTMC-reID数据集的Rank-1与mAP分别提高了3.68%和4.47%，这验证了相机风格转换后可以有效解决CVQStyle图像的风格差异问题；相比实验1，实验3加入了自校准卷积模块，Market-1501数据集的Rank-1与mAP分别提高了1.23%和1.92%，DukeMTMC-reID数据集的Rank-1与mAP分别提高了2.61%和3.58%，这验证了SCNet能够有效增大感受野，获取更多的特征信息；实验3结合了实验1和实验2的模块，实现了更好的效果，相比基础的Baseline，Market-1501数据集的Rank-1与mAP分别提高了3.50%和6.05%，DukeMTMC-reID数据集的Rank-1与mAP分别提高了5.10%和7.69%。

表2　不同模型的实验结果

Table 2　Experimental results of different models　　%

Experiment No.	Model	Market-1501		DukeMTMC-reID	
		Rank-1	mAP	Rank-1	mAP
1	Baseline	91.12	80.59	83.11	72.63
2	Baseline+CVQGAN	93.56	84.01	86.79	77.10
3	Baseline+SCNet	92.35	82.51	85.72	76.21
4	Baseline+ CVQGAN+ SCNet	94.62	86.64	88.21	80.32

3.5　实验结果可视化

本文将改进后模型的结果与基准模型的检索结果进行可视化展示，如图9所示。其中序号1～10为算法检索返回的相似度排名前10的样本图像，从左至右相似度逐次递减，图9(a)、(c)为基准模型的检索结果，图9(b)、(d)为本文模型的检索结果，矩形框代表错误的检索结果，即行人身份与查询结果不一致，如图9(a)中序号5,7,10对应的图片均为误检结果。对比结果表明，本文改进后的算法能够提取更多具有判别力的特征，有效地减小风格变化、视角变化等干扰信息的影响，这证明了本文所提算法的有效性。

3.6　与主流算法的比较

为了更直观地表明本文算法的有效性，在Market-1501和DukeMTMC-reID两种数据集上将本文算法与主流算法进行了比较。为了验证

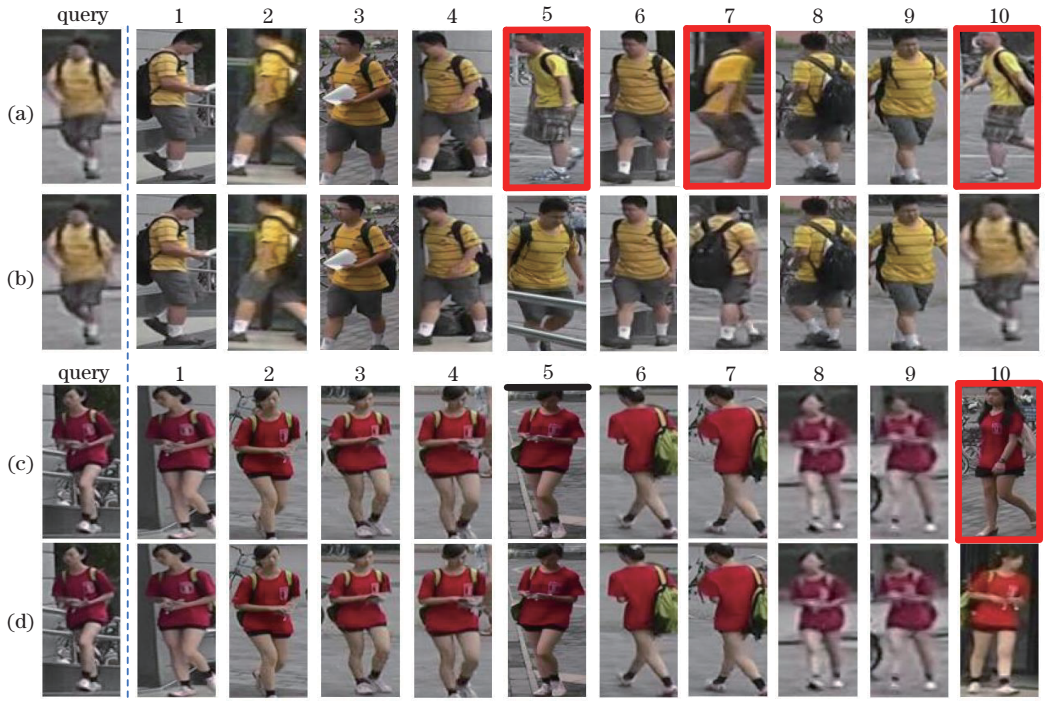

图 9　Market-1501 数据集的可视化结果。(a)(c)基准模型;(b)(d)所提模型

Fig. 9　Visualization results of Market-1501 dataset. (a) (c) Reference model; (b) (d) proposed model

CVQGAN 的广泛适用性，Baseline 除了使用 Resnet50 外，还将 PCB[24]、Densenet121[25]作为主干网络，将 CVQGAN 应用到以上两种模型，以验证其有效性，结果如表 3 所示。从表 3 可知，本文所提出的框架在两种数据集上的表现均超过了对比算法。

表 3　所提算法在 Market-1501 和 DukeMTMC-reID 数据集上与主流算法的性能比较

Table 3　Performance comparison of proposed algorithm with mainstream algorithms on Market-1501 and DukeMTMC-reID datasets　%

Algorithm	Market-1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
SVDNet[4]	82.30	62.10	76.70	56.80
Camstyle[16]	89.49	71.55	78.32	57.61
PAN[26]	82.81	63.35	71.59	51.51
LSRO[6]	83.97	66.07	67.68	47.13
PSE+ECN[27]	90.30	84.00	85.20	79.80
DCNN[28]	90.20	75.60	78.20	73.80
PCB[24]	92.40	77.30	81.90	65.30
DenseNet121[25]	90.17	76.02	80.62	63.32
PCB+CVQGAN	93.95	84.96	84.49	76.40
DenseNet121+CVQGAN	92.73	80.82	83.07	74.92
Proposed algorithm	94.62	86.64	88.21	80.32

4　结　　论

提出了一种基于 CVQGAN 与自校准卷积模块的行人重识别学习框架。通过给定任意摄像头下的一张行人图像，CVQGAN 将此样本图像转换为其他摄像头下清晰的、接近真实风格的不同行人图像，以此对数据集进行有效扩充，并且所提出的矢量量化模块有效解决了原始生成器产生噪声伪图像的问题，生成的 CVQStyle 图像质量更高。自校准卷积行人重识别网络将不同尺度的行人特征进行融合，从而获取更多的特征信息，使产生的特征图更具辨识性。所提方法在数据集 Market-1501 和 DukeMTMC-reID 上的性能与目前主流方法相比准确率和鲁棒性有了明显的提高，取得了更好的效果。

参　考　文　献

[1] 刘可文，房攀攀，熊红霞，等. 基于多层级特征的行人重识别[J]. 激光与光电子学进展，2020，57(8): 081503.
Liu K W, Fang P P, Xiong H X, et al. Person re-identification based on multi-layer feature[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081503.

[2] 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, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2288-2295.

[3] Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by local maximal occurrence representation and metric learning[C]//2015 IEEE Conference on Computer Vision and Pattern Recognition, June 7-12, 2015, Boston, MA, USA. New York: IEEE Press, 2015: 2197-2206.

[4] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3820-3828.

[5] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification[EB/OL]. (2017-03-22) [2021-06-02]. https://arxiv. org/abs/1703.07737.

[6] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline *in vitro*[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3774-3782.

[7] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. (2015-11-19)[2021-06-03]. https://arxiv.org/abs/1511.06434.

[8] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2242-2251.

[9] Yi Z L, Zhang H, Tan P, et al. DualGAN: unsupervised dual learning for image-to-image translation[C]//2017 IEEE International Conference on Computer Vision, October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 2868-2876.

[10] Kim T, Cha M, Kim H, et al. Learning to discover cross-domain relations with generative adversarial networks[C]//Proceedings of the 34th International Conference on Machine Learning, ICML 2017, August 6-11, 2017, Sydney, NSW, Australia. London: PMLR, 2017: 1857-1865.

[11] 徐龙壮, 彭力. 基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019, 56(14): 141504.
Xu L Z, Peng L. Person reidentification based on multiscale convolutional feature fusion[J]. Laser & Optoelectronics Progress, 2019, 56(14): 141504.

[12] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04) [2021-06-02]. https://arxiv. org/abs/1409.1556.

[13] 毕晓君, 汪灏. 基于视角信息嵌入的行人重识别[J]. 光学学报, 2019, 39(6): 0615007.
Bi X J, Wang H. Person re-identification based on view information embedding[J]. Acta Optica Sinica, 2019, 39(6): 0615007.

[14] Chen X S, Fu C M, Zhao Y, et al. Salience-guided cascaded suppression network for person re-identification[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 3297-3307.

[15] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial networks[EB/OL]. (2014-06-10)[2021-06-02]. https://arxiv.org/abs/1406.2661.

[16] Zhong Z, Zheng L, Zheng Z D, et al. Camera style adaptation for person re-identification[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 5157-5166.

[17] Kingma D P, Welling M. Auto-encoding variational Bayes[EB/OL]. (2013-12-20) [2021-06-03]. https://arxiv.org/abs/1312.6114.

[18] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[19] Liu J J, Hou Q B, Cheng M M, et al. Improving convolutional networks with self-calibrated convolutions[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 10093-10102.

[20] Zheng Z, Yang X, Yu Z, et al. Joint discriminative and generative learning for person re-identification [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 15-20, 2019, Long Beach, USA. New York: IEEE Press, 2019: 2138-2147.

[21] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark[C]//2015 IEEE

International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.

[22] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking[M]//Hua G, Jégou H. Computer vision-ECCV 2016 workshops. Lecture notes in computer science. Cham: Springer, 2016, 9914: 17-35.

[23] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. (2014-12-22) [2021-06-01]. https://arxiv.org/abs/1412.6980.

[24] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.

[25] Huang G, Liu Z, van der Maaten L, et al. Densely connected convolutional networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 2261-2269.

[26] Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(10): 3037-3045.

[27] Sarfraz M S, Schumann A, Eberle A, et al. A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 420-429.

[28] Li Y, Jiang X Y, Hwang J N. Effective person re-identification by self-attention model guided feature learning[J]. Knowledge-Based Systems, 2020, 187: 104832.