

基于迁移学习场景自 适应的车辆图像检索

本章开始研究车辆图像检索(Vehicle Re-Identification,简称车辆 ReID)技术。车辆图像检索技术是智能交通监控技术的重要组成部分,主要任务是给定一张查询车辆图像,在跨场景非重叠监控摄像机采集到的车辆图像中搜索具有相同身份(Identity,简称 ID)的目标车辆,并按照车辆的特征距离从小到大排序,从而判断监控摄像机是否采集到特定的目标车辆。

5.1 引言

随着深度学习在城市道路车辆检测任务中取得的广泛成功,一些基于卷积神经网络的车辆图像检索技术被提出,并在相同数据域的车辆图像中取得了较好的检索效果。但是,在实际应用中,不同数据域下的图像风格会有较大的变化,即用于训练的数据集与测试的数据集具有不同的图像风格。如图 5-1 所示, VeRi 数据集中的车辆图像清晰、光照充足,而 VRIC 数据集中的车辆图像则具有较低的分辨率且存在强烈的运动模糊。由于不同域中数据集之间的差异,导致在 VeRi 数据集上训练好的车辆图像检索模型在 VRIC 数据集上的性能表现不佳,反之亦然。



图 5-1 不同数据域中图像风格的对比

为了解决跨域场景下图像风格变化显著的问题,本章利用迁移学习的思想,将源域上的车辆图像迁移到目标域上,使迁移后车辆图像的风格与目标域图像的风格一致,同时迁移前后保持车辆的身份信息不变,将迁移后的车辆图像用于车辆检索任务,从而提升检索的效果。

目前,图像迁移学习的方法已经取得了巨大成功,例如,Mingsheng Long 等(2015)提出了一种新的深度自适应网络(Deep Adaptation Network,DAN)架构,将深度卷积神经网络推广到领域自适应场景,DAN 可以学习弥合跨域差异的可转移的特征,并通过内核嵌入的无偏估计进行线性扩展。Mingsheng Long 等(2016)提出了一种深度网络中域自适应的新方法,可以共同学习源域中标记数据和目标域中未标记数据的自适应分类器和可转移特征。Yuanwei Wu 等(2019)提出了一种简单有效的无监督深度特征转移算法,用于低分辨率图像分类。Yaël Frégier 等(2019)提出了一种使用 GAN 架构进行迁移学习的方法,并证明该方法训练时收敛更快,所需数据更少。近年来,迁移学习在行人图像检索领域得到了广泛的应用,例如,为了减轻标注新训练样本的昂贵成本,Longhui Wei 等(2017)提出了一个行人迁移生成对抗网络(Person Transfer Generative Adversarial Network,PTGAN)来弥合不同数据域之间的差异,实验表明,PTGAN 可以大大缩小域间的差距。Weijian Deng 等(2017)提出了相似性保持生成对抗网络(Similarity Preserving Generative Adversarial Network,SPGAN)实现跨域的行人图像检索,SPGAN 由孪生神经网络(Siamese Neural Network,SNN)和 CycleGAN 组成,通过领域适应性实验表明,SPGAN 生成的图像更适合于域自适应。Sergey Rodionov 等(2018)开发了一种在多个数据集上训练模型的方法,以及一种在线无监督微调的方法,将迁移学习应用到行人图像检索中。

因此,本章提出了一种基于迁移学习场景自适应的车辆图像检索(Transfer Learning Scenario Adaptation for Vehicle Re-Identification,TLSA)方法,如图 5-2 下面部分所示。TLSA 由两部分组成:①车辆迁移生成对抗网络(Vehicle Transfer Generative Adversarial Network,VTGAN);②基于图像风格迁移的车辆特征匹配与排序。具体地,给定一张源域中的查询车辆图像和目标域车辆图像数据库,首先,利用 VTGAN 将源域上的车辆图像迁移到目标域上,使迁移后的车辆图像具有目标域图像的风格;然后,提取迁移后车辆图像的深度特征形成 ReID 特征;最后,利用欧氏距离计算查询车辆图像与车辆图像数据库中两两图像 ReID 特征之间的相似度,并按照相似程度排序,实现有效的车辆图像检索。特别地,图 5-2 上面部分是非跨域场景下车辆图像检索的过程,网络直接提取查询车辆图像和车辆图像数据库中图像的深度特征,利用距离度量的方法计算其相似度,实现初步的车辆图像检索,因此,将其作为本章及本书后续其他方法对比的基准模型,即 Baseline。

本章设计了两组实验:①在 VeRi 数据集和 VRIC 数据集上对比了 3 种图像迁移方法,实验表明,VTGAN 生成的车辆图像具有较好的纹理细节,更接近目标域图像的风格。②在 VeRi 数据集和 VRIC 数据集上分别设计了多组对比实验和消融实验评估了 TLSA 车辆图像检索的效果,实验表明,当跨数据域时车辆图像检索的效果显著下降,而当采用迁移学习转换源域与目标域图像的风格后,车辆图像检索的性能有所提升,但仍与非跨域场景下的检索效果有较大的差距。

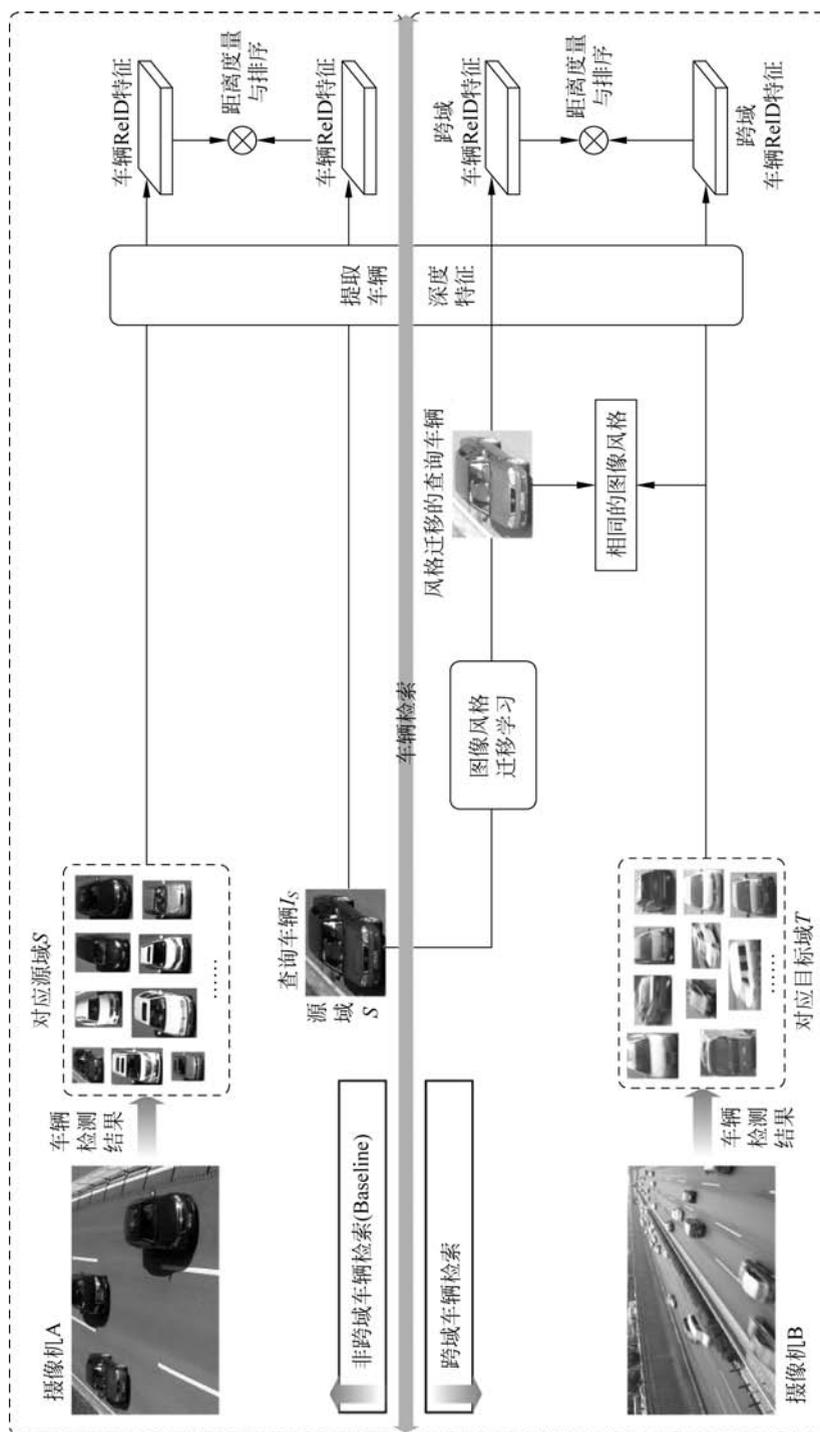


图 5-2 基于风格迁移的车辆图像检索模型框架

综上所述,本章的主要贡献有:①将连接-合并残差网络引入到车辆迁移生成对抗网络中,通过提升车辆特征的语义信息生成具有特定风格的车辆图像;②为保证迁移过程中车辆身份的一致性,利用孪生网络约束源域图像和目标域图像之间的映射关系;③通过转换车辆图像风格,提高了跨域场景下车辆图像检索的精度。

5.2 问题描述

根据前文中的定义,基于图像风格迁移的车辆图像检索可以描述为:给定一对车辆图像 (I_S, I_{T_i}) 以及它们相对应的二进制标签 (s_{ST_i}) ,其中 I_S 为源域 S 中的查询车辆图像, $I_{T_i} \in \{I_{T_i}\}_{i=1}^N$ 为目标域 T 中的图像。如果 I_S 与 I_{T_i} 是同一身份车辆在不同数据域中的图像,那么 $s_{ST_i}=1$;否则 $s_{ST_i}=0$,表示它们是不同的身份的车辆。对于源域 S 中的每一个查询图像 I_S ,我们的目标是将其映射成具有目标域 T 风格的特征向量 \mathbf{a} ,定义如下:

$$\mathbf{a} = F(I_{T_0}) \quad (5-1)$$

其中,

$$I_{T_0} = G(I_S, \{I_{T_i}\}_{i=1}^N) \quad (5-2)$$

这里, $\{I_{T_i}\}_{i=1}^N$ 为目标域车辆图像数据库, $G(\cdot)$ 表示生成器将输入的源域图像 I_S 迁移到目标域 T 中,使其具有 T 中图像 $\{I_{T_i}\}_{i=1}^N$ 的风格, I_{T_0} 为 I_S 迁移后生成的车辆图像, $F(\cdot)$ 表示获取图像的特征向量。此时,用于检索的车辆图像对表示为 (I_{T_0}, I_{T_i}) ,其中 $I_{T_i} \in \{I_{T_i}\}_{i=1}^N$,并且它们相对应的二进制标签更新为 $(s_{T_0T_i})$ 。

当车辆的ReID特征确定后,采用对比损失函数处理配对车辆图像的关系,其表达式如下:

$$L_{\text{reid}} = \frac{1}{2N} \sum_{i=1}^N [s_{T_0T_i} d^2 + (1 - s_{T_0T_i}) \max(m - d, 0)^2] \quad (5-3)$$

其中, N 为车辆图像的数量, $d = \|\mathbf{a}_{T_0} - \mathbf{a}_{T_i}\|_2$ 代表两个车辆特征向量 \mathbf{a}_{T_0} 和 \mathbf{a}_{T_i} 的距离度量, $s_{T_0T_i}$ 为两个车辆是否匹配的标签, $s_{T_0T_i}=1$ 表示两个车辆相似或者匹配, $s_{T_0T_i}=0$ 则表示不匹配, m 为设定的阈值。

从对比损失函数的表达式(5-3)可以发现,该损失函数可以很好地表达对车辆图像的匹配程度。

(1) 当 $s_{T_0T_i}=1$ 时,表明两个车辆相似或匹配,其损失函数表示为 $L_{\text{reid}} = \frac{1}{2N} \sum_{n=1}^N s_{T_0T_i} d^2$ 。若两个相似车辆在特征空间的距离较大,则损失值会变大,函数会对该正样本对进行惩罚,从而缩短特征向量 \mathbf{a}_{T_0} 和 \mathbf{a}_{T_i} 之间的距离,如图5-3所示。

(2) 当 $s_{T_0T_i}=0$ 时,表明两个车辆不相似或不匹配,其损失函数表示为 $L_{\text{reid}} =$

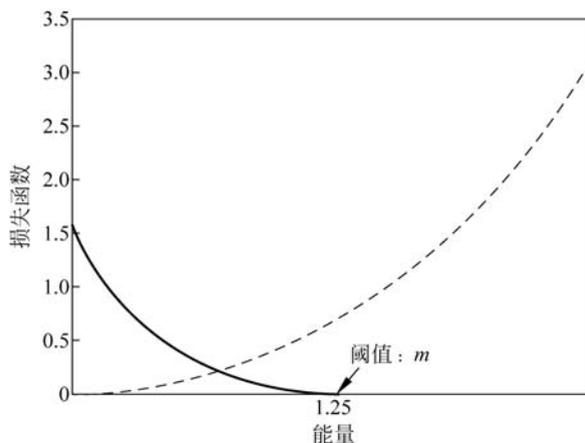


图 5-3 对比损失函数值与车辆特征的欧氏距离之间的关系

$\frac{1}{2N} \sum_{i=1}^N (1 - s_{T_0 T_i}) \max(m - d, 0)^2$ 。若两个不相似车辆在特征空间的距离较小,则损失值会增大,函数会对该负样本对进行惩罚,从而增加特征向量 \mathbf{a}_{T_0} 和 \mathbf{a}_{T_i} 之间的距离使其大于阈值 m ,如图 5-3 所示。

当模型训练完成后,计算源域 S 中的查询车辆图像和目标域 T 中的车辆图像数据库两两特征向量之间的欧氏距离,并按照相似程度进行排序。

5.3 车辆迁移生成对抗网络

为了使在源域 S 上训练的模型在目标域 T 上同样获得较好的车辆图像检索效果,提出了一种车辆图像风格迁移模型,将源域 S 上的车辆图像迁移到目标域 T 上,使迁移后车辆图像的风格与目标域 T 的风格一致,同时保持迁移前后车辆的身份信息不变,将迁移后的车辆图像用于车辆检索模型的训练,该迁移模型称为车辆迁移生成对抗网络 (Vehicle Transfer Generative Adversarial Network, VTGAN)。如图 5-4 所示,VTGAN 由两个生成器网络 (G 和 F)、两个判别器网络 (D_T 和 D_S),以及一个孪生网络 SiaNet 组成。首先,在源域 S 上获取输入的车辆图像,该输入图像被传递到第一个生成器网络,表示为 $G: S \rightarrow T$,其任务是将源域 S 上给定的车辆图像转换到目标域 T 的图像中。然后,新生成的车辆图像被传递到另一个生成器网络,表示为 $F: T \rightarrow S$,其任务是将生成的图像转换回源域 S 中的图像,两个域的训练集图像不成对出现,且转换后的图像分别被送入对应的判别器 D_T 和 D_S 做训练,使转换回的图像与原始输入的图像不可被分辨,此时生成对抗网络的结构可定义为循环的形式,即 $F(G(S)) \approx S$ 。最后,孪生网络 SiaNet 用于确保迁移前后车辆的身份信息保持不变。

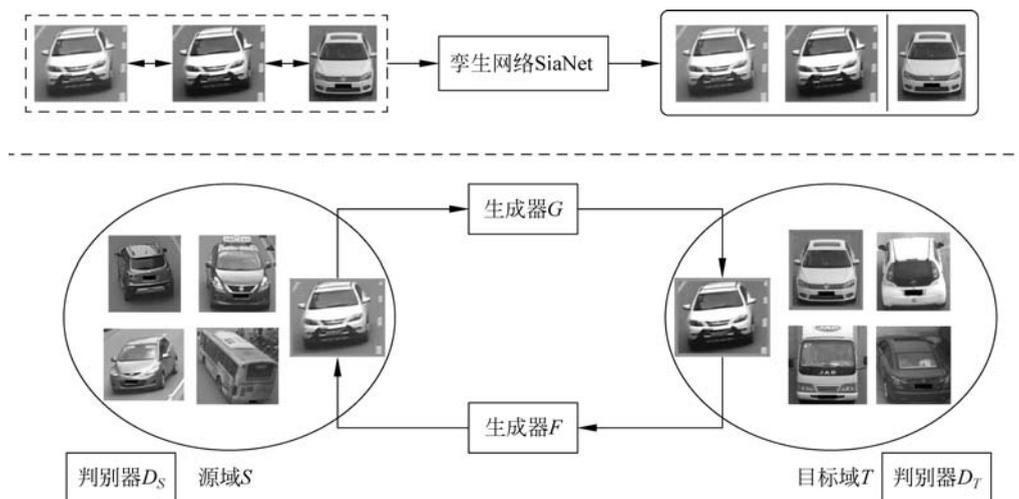


图 5-4 车辆迁移生成对抗网络的模型结构

5.3.1 生成器网络

用于车辆图像迁移的生成器网络 G 和 F 的网络结构一样,均由编码器、转换器和解码器三部分构成,如图 5-5 所示。借鉴 Pix2Pix 和 CycleGAN 转换两个不同域中图像风格的方法,生成器网络中输入不成对出现的车辆图像,并缩放为 256×256 分辨率。首先,编码器利用 3 个卷积层提取车辆特征,将图像压缩成 256 个 64×64 的特征向量;然后,转换器使用 9 个如图 4-4(c)所示的连接-合并残差块学习组合车辆图像的不相近特征,将图像在源域 S 中的特征向量转换为目标域 T 中的特征向量,并在转换的同时保留原始图像的身份信息;最后,解码器利用 2 个反卷积层、1 个卷积层,以及 Tanh 激活函数从特征向量中恢复图像的低级特征,得到图像风格转换的车辆图像。

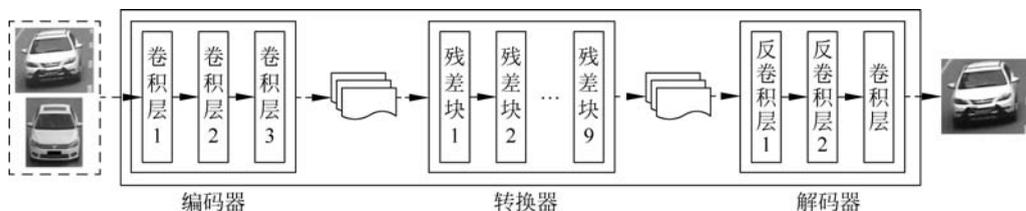


图 5-5 车辆迁移生成器网络的结构

5.3.2 判别器网络

生成器网络 G 和 F 对应的判别器网络分别为 D_T 和 D_S ,同样 D_T 和 D_S 的网络结构是

一样的。如图 5-6 所示,给定生成器网络输出的车辆图像,判别器网络首先使用 4 个卷积层提取图像特征,得到 512 个 32×32 的特征向量,然后添加一维输出的卷积层确定提取的特征是否属于特定的类别。

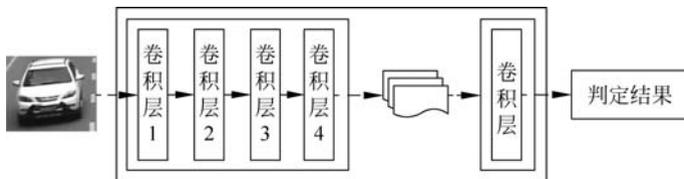


图 5-6 车辆迁移判别器网络的结构

5.3.3 孪生网络

车辆图像风格迁移的核心任务是:要保持图像迁移前后车辆身份相关的信息不变,这种身份信息不是图像的背景或图像的风格,而是和身份信息有潜在关系的车辆区域。前面的生成器网络与判别器网络学习了源域 S 与目标域 T 之间的映射关系,因此,本节利用孪生神经网络 SiaNet 约束学习到的映射关系,使迁移前后同一车辆图像之间的特征距离近一些,而不同车辆图像之间的特征距离远一些。具体地,孪生网络 SiaNet 由两个结构相同、权值共享的卷积神经网络组成,每个卷积神经网络独立地提取图像特征,分别由 4 个卷积层、4 个最大池化层和 1 个全连接层构成,如表 5-1 所示,在网络的最后一层定义了对比损失函数评价两个输入车辆图像的相似度。

表 5-1 单个孪生模型的网络结构

| 类 型 | 卷 积 核 | 步 长 | 输 出 |
|-------|--------------|-----|------------------|
| 卷积层 | 4×4 | 2 | 64×64 |
| 最大池化层 | 2×2 | 2 | 32×32 |
| 卷积层 | 4×4 | 2 | 128×128 |
| 最大池化层 | 2×2 | 2 | 64×64 |
| 卷积层 | 4×4 | 2 | 256×256 |
| 最大池化层 | 2×2 | 2 | 128×128 |
| 卷积层 | 4×4 | 2 | 512×512 |
| 最大池化层 | 2×2 | 2 | 256×256 |
| 全连接层 | — | — | 128×128 |

5.3.4 网络训练

生成对抗网络旨在使用最小-最大的策略在训练数据和生成样本之间找到最佳的判别器网络 D ,同时增强生成器网络 G 的性能,根据交叉熵损失,可以构造下面的损失函数:

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}(x)}} [\log D(x)] + E_{z \sim P_z(z)} [\log(1 - D(G(z)))] \quad (5-4)$$

其中, $P_{\text{data}(x)}$ 是真实样本的分布, $P_{z(z)}$ 是服从高斯分布的噪声数据, 训练过程中, 生成器网络 G 的目标是尽量生成逼真的图像, 而判别器网络 D 的目标是尽量辨别出 G 生成的假图像和真实的图像。

VTGAN 由对抗损失、循环一致损失、身份映射损失以及对比损失等多个损失函数构成。首先, VTGAN 引入两个生成器-判别器对: $\{G, D_T\}$ 和 $\{F, D_S\}$, 它们分别将源域 S 中的车辆图像映射到目标域 T , 或将目标域 T 中的车辆图像映射到源域 S 。对于生成器 G 和与其对应的判别器 D_T , 定义对抗损失函数为:

$$L_T(G, D_T, p_x, p_y) = E_{y \sim p(y)} [(D_T(y) - 1)^2] + E_{x \sim p(x)} [(1 - D_T(G(x)))^2] \quad (5-5)$$

其中, p_x 和 p_y 分别表示源域 S 和目标域 T 中车辆图像的分布情况。对于生成器 F 和与其对应的判别器 D_S , 对抗损失函数定义如下:

$$L_S(F, D_S, p_y, p_x) = E_{x \sim p(x)} [(D_S(x) - 1)^2] + E_{y \sim p(y)} [(1 - D_S(F(y)))^2] \quad (5-6)$$

由于训练数据是不成对出现的, 源域 S 与目标域 T 之间存在无限个映射函数, 因此, 引入循环一致损失函数来减少映射函数可能的表达空间。循环一致损失函数表示如下:

$$L_{\text{cyc}}(G, F) = E_{x \sim p(x)} [\|F(G(x)) - x\|_1] + E_{y \sim p(y)} [\|G(F(y)) - y\|_1] \quad (5-7)$$

此外, 引入目标域身份映射损失约束源域 S 与目标域 T 之间映射关系的学习, 尽可能保证转换前后车辆图像的相似性。目标域身份映射损失函数为

$$L_{\text{ide}}(G, F, p_x, p_y) = E_{x \sim p(x)} \|F(x) - x\|_1 + E_{y \sim p(y)} \|G(y) - y\|_1 \quad (5-8)$$

同时, 实验表明, 该目标域身份映射损失函数能够使迁移前后图像的颜色保持一致, 如图 5-7 所示, 与第三行相比, 第二行为去掉 L_{ide} 损失的效果。为了进一步保留车辆的身份信息, 孪生网络 SiaNet 利用对比损失函数控制正负样本的距离, 对比损失函数表示如下:

$$L_{\text{con}}(i, x_1, x_2) = (1 - i) \{\max(0, m - d)\}^2 + id^2 \quad (5-9)$$

$$d(x_1, x_2) = \|x_1 - x_2\|_2 \quad (5-10)$$



图 5-7 不同方法迁移车辆图像风格的结果比较

其中, x_1 和 x_2 为输入的图像样本对, d 代表两个输入图像特征向量之间的欧氏距离, $m \in [0, 2]$ 为设定的阈值, i 为输入图像对的标签, 当输入图像 x_1 和 x_2 为正样本对时, $i=1$, 否则, $i=0$ 。

因此, 车辆迁移生成对抗网络(VTGAN)最终的损失函数由以上对抗损失、循环一致损失、身份映射损失和对比损失构成, 并联合训练, 表示如下:

$$L_{\text{vtgan}} = L_T + L_S + \lambda_1 L_{\text{cyc}} + \lambda_2 L_{\text{ide}} + \lambda_3 L_{\text{con}} \quad (5-11)$$

其中, λ_1 、 λ_2 和 λ_3 控制三个损失函数的重要程度, 前三个损失函数 L_T 、 L_S 和 L_{cyc} 属于 CycleGAN 的方法, 身份映射损失 L_{ide} 以及对比损失 L_{con} 为图像迁移过程中增加的具体约束。

5.4 基于图像风格迁移的车辆图像检索

如图 5-2 所示, 跨域场景下的车辆图像检索包括图像风格迁移和特征匹配两个过程, 本节关注车辆特征学习和损失函数的设计。

5.4.1 特征学习

为获得车辆图像丰富的语义信息, 采用深度残差网络提取车辆特征, 该网络在 ResNet-50 的基础上, 修改残差块的堆叠方式, 利用第 4 章提出的连接-合并残差块替换 ResNet-50 中传统的残差块, 形成基于连接-合并残差块的特征提取网络, 记为 CMNet-50, 其中连接-合并残差块的结构如图 4-4(c) 所示。CMNet-50 在 ImageNet-2012 数据集上进行预训练, 并在给定的车辆图像检索数据集上进行微调。在测试阶段, 给定输入车辆图像, 提取 CMNet-50 中卷积层 5 的特征获得图像的描述子, 并采用欧氏距离进行车辆匹配。

为进一步提升在目标域中车辆图像检索的性能, 本节引入局部最大池化(Local Max Pooling, LMP)方法, 替代原始 CMNet-50 卷积层 5 后面的全局平均池化(Global Average Pooling, GAP)。如图 5-8 所示, 首先将卷积层 5 的特征图水平地划分成两部分, 然后在每

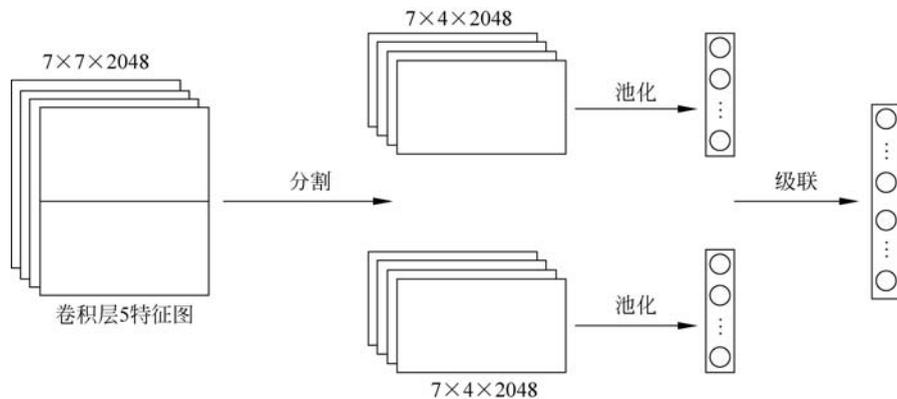


图 5-8 局部最大池化示意图

一部分上分别执行 GAP,最后融合每一部分的 GAP 输出形成 LMP 作为最终的特征表示,该网络记为 CMNet-50-LMP。实验表明,引入局部最大池化能够在一定程度上消除噪声数据带来的负面影响。

5.4.2 损失函数

本章采用基于身份学习的车辆图像检索方法。假设有 N 张车辆图像,包含 K 个不同的身份。令 $D_i = \{x_i, d_i\}$ 作为训练集,其中, $i \in N$, x_i 表示第 i 张车辆图像, d_i 表示图像 x_i 的身份标签。

给定训练样本 x ,利用 CMNet-50-LMP 网络获取其 ReID 特征,记为 f ,输出向量的尺寸为 $1 \times 1 \times 2048$,随后全连接层的输出为 $z = [z_1, z_2, \dots, z_K] \in \mathbb{R}^K$ 。预测每个车辆身份的概率计算如下:

$$p(p | x) = \frac{\exp(z_k)}{\sum_{i=1}^K \exp(z_i)} \quad (5-12)$$

因此,车辆身份分类的交叉熵损失函数计算如下:

$$L_{\text{ID}}(f, d) = - \sum_{k=1}^K \log(p(k))q(k) \quad (5-13)$$

令 y 为正确的车辆身份标签,对于所有 $k \neq y$,使 $q(y) = 1$ 且 $q(k) = 0$,即最小化交叉熵损失等同于最大化获得正确车辆身份类别的可能性。

本书第 5 章、第 6 章和第 7 章的车辆图像检索任务均使用以上基于身份的交叉熵损失函数训练车辆检索模型,不同之处在于其 ReID 特征 f 不同。

5.4.3 基于风格迁移的车辆图像检索

如图 5-2 所示,上面部分为基准的车辆图像检索框架(即 Baseline),下面部分为基于车辆图像风格迁移的车辆检索框架。给定源域 S 中的查询车辆图像 I_S 和目标域数据集 T ,图像检索实现过程如下:

(1) 将源域 S 中的查询车辆 I_S 的图像风格转换成目标域 T 的风格,记为 I_{T_0} ,此时,查询车辆图像 I_{T_0} 与目标域数据集 T 具有相同的图像风格。

(2) 利用特征提取网络 CMNet-50-LMP 分别获得车辆图像 I_{T_0} 与目标域数据集 T 中每一个车辆图像的 ReID 特征,分别记作 f_{T_0} 和 $f_{\{T_i\}_{i=1}^N}$,其中 N 为车辆的数量。

(3) 使用欧氏距离计算源域中查询车辆图像 I_{T_0} 与目标域中车辆图像数据库 T 中两两图像 ReID 特征之间的相似度,其距离公式表示为

$$d_{I_q, \{I_{g_i}\}_{i=1}^N} = \|f_q - f_{\{g_i\}_{i=1}^N}\|_2 \quad (5-14)$$

通过最小化式(5-3)中的对比损失函数 L_{reid} ,使相同车辆图像(正样本对)的距离尽可能小,不同车辆图像(负样本对)的距离尽可能大,最终实现跨域场景下的车辆图像检索。

5.5 实验结果与分析

首先评估车辆迁移生成对抗网络(VTGAN)转换图像风格的性能,然后研究基于VTGAN的车辆图像检索方法 TLSA 在跨域应用中的表现,最后评估基于局部最大池化的特征提取网络对 TLSA 性能的影响。本章的实验(包括 VTGAN 和 TLSA 两部分)均基于 PyTorch 网络框架实现,并在配置有 Intel Core i7-7700K CPU 和 NVIDIA GTX 1080Ti GPU 的 PC 上运行。

5.5.1 数据集

为验证图像风格迁移对车辆图像检索效果的影响,使用 VeRi 数据集和 VRIC 数据集进行实验。由于 VRIC 数据集从真实场景的 UA-DETRAC 训练视频中裁剪获得,具有真实的多分辨率、运动模糊、照明变化、遮挡和不同视角等特点,同时与 VeRi 数据集风格差异较大,因此,可以更好地反映实验效果。

1. VeRi 数据集

VeRi 数据集由北京邮电大学收集、标注并发布。VeRi 数据集由部署在约 1km^2 城市区域内的 20 个监控摄像机采集获得,摄像头安装角度包括正视角、侧视角与斜视角等,同时拍摄多个不同的交通场景,如十字路口、转弯路口、单向车道、双向车道、二车道道路、四车道道路等。VeRi 数据集不仅标注了车辆的身份 ID、采集摄像机 ID,还标注了车辆的颜色属性,如 yellow、orange、green、gray、red、blue、white、golden、brown、black,以及车辆的类型,如 sedan、suv、van、hatchback、mpv、pickup、bus、truck、estate。

VeRi 数据集包含 776 个不同身份车辆的 49357 张图像,其中 576 个不同身份车辆的 37778 张图像用于训练,其余 200 个不同身份车辆的 11579 张图像用于测试,同时从测试集中随机选取 1678 张车辆图像作为待查询图像。VeRi 数据集中的车辆图像包含丰富的车辆视角、颜色、类型、尺寸,以及不同的光照条件与场景,如图 5-9 所示。

2. VRIC 数据集

VRIC 数据集由伦敦玛丽女王大学电子工程和计算机科学学院发布。现有的车辆图像检索评价基准严重地依赖高质量的图像以及其恒定尺寸的车辆细粒度的外观,这类似于车牌识别中对图像质量的要求,而在现实的车辆图像检索场景中,车辆图像的分辨率是任意改变的。为在实际场景中验证方法的有效性,引入更具有挑战性的车辆图像检索数据集基准——VRIC 数据集。VRIC 数据集从 60 个带有边界框注释的 UA-DETRAC 训练视频中获取,UA-DETRAC 训练集由 60 个不同的摄像机在白天和夜间的异构道路交通场景中捕获。为确保足够的车辆外观变化,VRIC 数据集丢弃出现轨迹小于 20 帧以及边界框小于 24×24 分辨率的车辆图像。



图 5-9 VeRi 数据集中的车辆图像样本

VRIC 数据集包含 5622 个不同身份车辆的 60430 张图像,其中 54808 张车辆图像用于训练,2811 张车辆图像用于测试,另外 2811 张车辆图像作为待查询图像。如图 5-10 所示,与现有车辆图像检索数据集相比,VRIC 数据集最大的特点在于车辆图像的分辨率及比例变化多样,同时在运动模糊、照明、遮挡和视角等方面更逼真。



(a) 不同气候条件下的UA-DETRAC数据集,如多云、夜晚、晴天、雨天



(b) VRIC数据集中的车辆图像样本

图 5-10 UA-DETRAC 数据集与 VRIC 数据集中的车辆图像样本

表 5-2 显示了两个数据集中数据的统计和分布情况。

表 5-2 两个数据集中的数据统计和分布

| 名称 | 类型 | 总计 | 训练集 | 测试集 | |
|----------|--------|-------|-------|------|-------|
| | | | | 待查询 | 检索库 |
| VeRi 数据集 | 车辆身份数目 | 776 | 576 | 200 | 200 |
| | 车辆图像数目 | 49357 | 37778 | 1678 | 11579 |
| VRIC 数据集 | 车辆身份数目 | 5622 | 2811 | 2811 | 2811 |
| | 车辆图像数目 | 60430 | 54808 | 2811 | 2811 |

5.5.2 评价指标与实验设置

车辆图像检索任务中,给定一个查询车辆图像,期待在车辆图像数据库中找到具有相同身份的候选图像,并按照相似程度的高低进行排序。

1. 评价指标

受到行人图像检索研究的启发,车辆图像检索任务的准确率采用排序第一准确率(Rank@1)、排序第五准确率(Rank@5)、排序第二十准确率(Rank@20)以及平均精度均值(mean Average Precision, mAP)进行评价。

Rank@k 表示给定查询车辆图像,在车辆图像数据库中进行搜索,搜索结果按照相似度排序后的前 k 张图像中存在与查询图像属于同一身份的准确率。假设车辆身份的数量为 N ,利用训练好的模型提取查询图像的特征,如下:

$$\mathbf{f}_{\text{Query}} = [\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_N] \quad (5-15)$$

车辆图像数据库含有 M 张图像,提取其中所有图像的特征如下:

$$\mathbf{f}_{\text{Gallery}} = \begin{bmatrix} \mathbf{a}_{11} & \mathbf{a}_{12} & \dots & \mathbf{a}_{1N} \\ \mathbf{a}_{21} & \mathbf{a}_{22} & \dots & \mathbf{a}_{2N} \\ \vdots & \vdots & & \vdots \\ \mathbf{a}_{M1} & \mathbf{a}_{M2} & \dots & \mathbf{a}_{MN} \end{bmatrix} \quad (5-16)$$

利用欧氏距离计算图像特征之间的相似度 S ,表示如下:

$$S_M = \sqrt{\sum_{k=1}^N (\mathbf{q}_k - \mathbf{a}_{Mk})^2} \quad (5-17)$$

其中, \mathbf{q}_k 为一个查询图像的特征向量, \mathbf{a}_{Mk} 为车辆图像数据库中第 M 张图像的特征向量。 S_M 越低表示图像之间的相似性越大,即是同一个车辆的置信度越高。因此,将 S_M 按照大小从低到高排序,分别计算 Rank@k, $k=1, 5, 20$ 的准确率。

对于每一个查询车辆图像,当车辆图像数据库中有多个目标图像与其匹配时,使用均值平均精度(mAP)平衡模型的查准率(Precision)与召回率(Recall),从而评估模型的整体性能。具体地,给定一个查询车辆图像,平均精度(Average Precision, AP)可表示为

$$AP = \frac{\sum_{k=1}^M \text{Rank}@k \times \text{gt}(k)}{N_{\text{gt}}} \quad (5-18)$$

其中, N_{gt} 为查询车辆图像对应的正样本数量, $\text{gt}(k)$ 为标记函数, 当数据库中第 k 个图像与查询车辆图像匹配时返回 1, 否则返回 0。对于全部查询车辆图像, 其平均精度均值(mAP)表示如下:

$$\text{mAP} = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (5-19)$$

其中, Q 为全部查询车辆图像的数量。

2. 实验设置

对于图像迁移模型 VTGAN, 利用 VeRi 数据集和 VRIC 数据集的训练图像进行训练。在所有实验中, 设定式(5-9)中的参数 $m=2$, 式(5-11)中的参数 $\lambda_1=10$, $\lambda_2=5$, $\lambda_3=2$ 。训练 VTGAN 时, 将输入图像大小裁剪为 256×256 , 利用 Adam 算法(如算法 5-1 所示)优化网络参数。整个网络训练 200 个 epoch, batch_size 设置为 4, 初始学习率(learning_rate)为 0.0002。测试时, 使用生成器 G 将 VeRi 图像转换为 VRIC 图像风格, 使用生成器 F 将 VRIC 图像转换为 VeRi 图像风格, 转换后的图像用于微调在源域图像上训练的模型。

对于模型 TLSA, 将所有训练图像的大小调整为 256×256 , 并从所有图像中减去由训练图像计算获得的均值。在训练过程中, 将图像的大小随机裁剪为 224×224 , 对数据集进行随机抽样并使用图像的随机顺序。整个网络训练 60 个 epoch, batch_size 设置为 64, 初始学习率(learning_rate)为 0.05, 且每迭代 20 个 epoch 后学习率以 0.1 倍递减。使用小批量的随机梯度下降算法 SGD(如算法 4-1 所示)更新网络参数, 在网络最后的卷积层之前使用 dropout 函数, 并将其参数设置为 0.5。

算法 5-1 Adam 算法

Require: 步长 ϵ

Require: 矩估计的指数衰减速度, ρ_1 和 ρ_2 在区间 $[0, 1]$ 内

Require: 用于数值稳定的小常数 δ

Require: 初始参数 θ

初始化一阶和二阶矩变量 $s=0, r=0$

初始化时间步 $t=0$

while 没有达到停止准则 do

 从训练集中采集包含 m 个样本 $\{x^{(1)}, \dots, x^{(m)}\}$ 的小批量, 对应目标为 $y^{(i)}$

 计算梯度估计: $g \leftarrow \frac{1}{m} \nabla_{\theta} \sum_i L(f(x^{(i)}; \theta), y^{(i)})$

$t \leftarrow t+1$

 更新有偏一阶矩估计: $s \leftarrow \rho_1 s + (1-\rho_1)g$

 更新有偏二阶矩估计: $r \leftarrow \rho_2 r + (1-\rho_2)g \odot g$

$$\text{修正一阶矩的偏差: } \hat{s} \leftarrow \frac{s}{1-\rho_1^t}$$

$$\text{修正二阶矩的偏差: } \hat{r} \leftarrow \frac{r}{1-\rho_2^t}$$

$$\text{计算更新: } \Delta\theta = -\epsilon \frac{\hat{s}}{\sqrt{\hat{r}} + \delta}$$

$$\text{应用更新: } \theta \leftarrow \theta + \Delta\theta$$

end while

图 5-11 显示了 TLSA 在 VeRi 数据集和 VRIC 数据集上训练过程的 Loss(损失)曲线, 由于学习率的变化, Loss 曲线在第 20 个 epoch 的时候有震荡, 随后很快收敛。

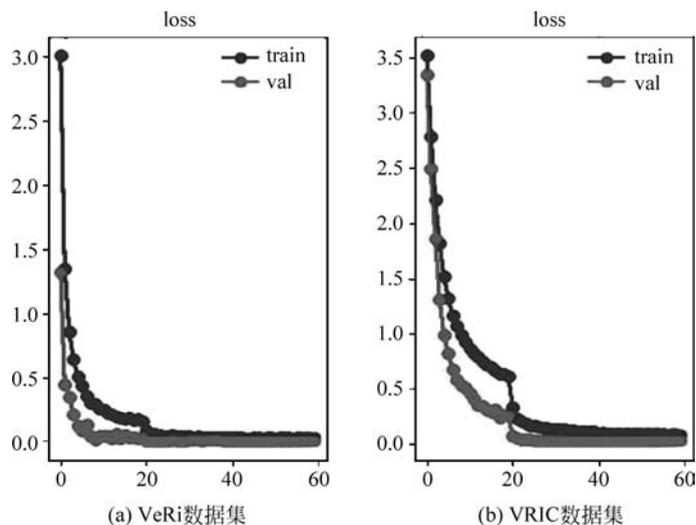


图 5-11 TLSA 在两个目标域上训练的 Loss 曲线

5.5.3 车辆图像风格迁移评估

为了验证车辆迁移生成对抗网络(VTGAN)转换图像风格的效果, 在 VeRi 数据集和 VRIC 数据集上对 3 种迁移方法进行了比较, 具体如下:

方法一: 基准方法(CycleGAN)。CycleGAN 是目前典型的图像风格迁移的方法, 能够在没有成对训练数据的情况下, 将图像内容从源域迁移到目标域。在训练时, CycleGAN 只需要将源域的图像和目标域的图像作为输入即可, 不要求源域跟目标域的图像内容匹配。

方法二: 目标域身份映射损失(CycleGAN+L_{ide})。该方法在 CycleGAN 的基础上增加了目标域身份映射损失函数 L_{ide} , 该损失函数可以保证迁移前后车辆图像的相似性和颜色的一致性。

方法三：对比损失(VTGAN)。该方法为本章提出的车辆迁移生成对抗网络,在方法二的基础上增加对比损失函数 L_{con} ,通过学习一种映射关系,使相同车辆之间的距离变小、不同车辆之间的距离变大;同时修改特征提取网络,使用连接-合并残差网络(CMNet-50)与局部最大池化(LMP)相结合的结构来获得更有效的车辆信息。

图 5-7 显示了使用以上 3 种方法在 VeRi 数据集与 VRIC 数据集之间转换图像风格获得的车辆图像效果,从图中可以得出如下结论:

(1) 将 VeRi 数据集转换为 VRIC 数据集风格,获得的车辆图像光照变暗、具有运动模糊的效果,与 VRIC 数据集的风格一致。

(2) 将 VRIC 数据集转化为 VeRi 数据集风格,获得的车辆图像轮廓清晰、具有较好的光照强度和色彩饱和度,与 VeRi 数据集的风格相近。

(3) 使用 CycleGAN 生成的图像具有完整的车辆轮廓,但图像清晰度不高、部分车辆信息丢失。

(4) 在 CycleGAN 的基础上增加目标域身份映射损失后,生成的车辆图像保留了更丰富的信息,图像更加完整,色彩更加真实,但图像仍然模糊。

(5) 增加对比损失函数并增强特征提取网络,使 VTGAN 生成的车辆图像更加完整与清晰、具有较好的纹理细节,更接近目标域图像的风格。

另外,图 5-12 展示了使用 VTGAN 将 VeRi 数据集迁移到 VRIC 图像风格的结果示例。可以清楚地看到, VeRi 图像的分辨率有较大的变化,同时具有了运动模糊的效果,说明 VTGAN 生成的车辆图像具有目标域的风格,并保留了源域中车辆的身份信息。



图 5-12 VeRi 中的车辆图像迁移到 VRIC 风格的结果

图 5-13 展示了使用 VTGAN 将 VRIC 数据集迁移到 VeRi 图像风格的结果示例。可以发现,VRIC 图像的清晰度与饱和度都有了变化,一定程度上具有了 VeRi 图像的风格,说明深层的 VTGAN 网络可以学习到特定任务的图像特征,并完成图像特征之间的迁移。



图 5-13 VRIC 中的车辆图像迁移到 VeRi 风格的结果

5.5.4 车辆图像检索性能评估

针对车辆图像检索任务,由于数据集之间的差异,在源域 S 上训练的模型在目标域 T 上的性能表现会剧烈下降,因此,本章提出利用 VTGAN 转换源域 S 与标域 T 之间的图像风格,通过迁移学习提高车辆图像检索的结果。本节从迁移学习和特征提取等角度分析 TLSA 方法的有效性。

1. 数据集分配明细

本节将对比基于不同训练数据的车辆图像检索结果,主要包括 3 类方案:非图像迁移的方法、图像域直接交换的方法以及图像迁移的方法。以源域 VRIC 数据集和目标域 VeRi 数据集为例,非图像迁移的方法是指目标域 VeRi 数据集分别作为训练集与测试集,该类方法为本文参考对比的 Baseline 方法;图像域直接交换的方法是指源域 VRIC 作为训练集、目标域 VeRi 作为测试集,该类方法用于验证车辆直接跨域检索的效果;图像迁移的方法是指源域 VRIC 转换成目标域 VeRi 风格的图像作为训练集、目标域 VeRi 作为测试集,该类方法用于本章提出的基于迁移学习的车辆图像检索方法(TLSA)。

表 5-3 显示了本章实验中数据集的分配情况。其中, G 为 VTGAN 的生成器, $G(\text{VRIC})$ 表示将源域 VRIC 数据集中的图像迁移转换为目标域 VeRi 风格的图像, $G(\text{VeRi})$ 表示将源域 VeRi 数据集中的图像迁移转换为目标域 VRIC 风格的图像。

表 5-3 基于迁移学习车辆图像检索的实验数据集明细

| 方 法 | VeRi 数据集 源域: VRIC, 目标域: VeRi | | VRIC 数据集 源域: VeRi, 目标域: VRIC | |
|-----------------|---------------------------------|------|---------------------------------|------|
| | 训练集 | 测试集 | 训练集 | 测试集 |
| Baseline | VeRi | VeRi | VRIC | VRIC |
| Direct Transfer | VRIC | VeRi | VeRi | VRIC |
| TLSA | $G(\text{VRIC})$ | VeRi | $G(\text{VeRi})$ | VRIC |

2. TLSA 性能评估

1) 直接转换与基准方法的比较

在表 5-4 中对比基准方法(Baseline)和直接跨域的图像检索方法(Direct Transfer),可以清楚地看到,车辆检索模型在跨数据域时的性能显著下降。具体地,利用 CMNet-50-LMP 特征提取网络,在目标域 VeRi 上训练的模型在目标域 VeRi 上测试的时候,mAP 为 60.01%,但是在源域 VRIC 上训练的模型在目标域 VeRi 上测试的时候,mAP 降低为 10.11%。当 VRIC 数据集作为目标域时,可以看到模型的性能同样会大幅度下降,这是因为不同域中的车辆图像之间具有较大的差异。

表 5-4 在目标域上不同方法性能的比较(%)

| 方 法 | VeRi 数据集 | | | | VRIC 数据集 | | | |
|-----------------|----------|--------|--------|---------|----------|--------|--------|---------|
| | mAP | Rank@1 | Rank@5 | Rank@20 | mAP | Rank@1 | Rank@5 | Rank@20 |
| Baseline | 60.01 | 89.63 | 95.47 | 96.84 | 64.13 | 60.05 | 83.26 | 90.62 |
| Direct Transfer | 10.11 | 29.50 | 41.95 | 61.26 | 20.14 | 16.72 | 29.63 | 45.00 |
| TLSA | 23.75 | 44.26 | 52.60 | 75.86 | 32.57 | 34.18 | 48.69 | 59.53 |

2) TLSA 方法的效果

通过对比 Direct Transfer(在源域 S 上训练的模型直接用于目标域 T)发现,本章提出的 TLSA 在 VeRi 数据集上的 mAP 和 Rank@1 精度分别提升了 13.64%和 14.76%,同样在 VRIC 数据集上分别有 12.43%和 17.46%的提升,这说明本章提出的 TLSA 对于跨域场景下的车辆图像检索任务是有效的。

图 5-14 展示了 TLSA 方法在目标域 VeRi 数据集上的车辆图像检索结果。图中每一



图 5-14 TLSA 在 VeRi 数据集上的车辆图像检索结果示例

行的最左侧为输入的查询车辆图像,右侧列出了检索结果中排序前 20 的车辆图像。在检索结果的图像上面显示了排序编号,其中带方框的编号为正确的检索结果、不带方框的编号为错误的检索结果。从图中可以看到,使用具有 VeRi 风格的 VRIC 数据集训练的 TLSA 模型,在不同程度上可以检索到正确的车辆,如图中带有方框编号的结果,但检索正确车辆的数目不多,这符合车辆跨域、跨摄像机追踪的事实。

图 5-15 展示了 TLSA 方法在目标域 VRIC 数据集上的车辆图像检索结果。从图中看到,虽然车辆检索的效果显著降低,但是对于分辨率变化较大的 VRIC 数据集,在排序前 10 的结果中都有检索正确的车辆,说明 TLSA 方法具有较好的鲁棒性。

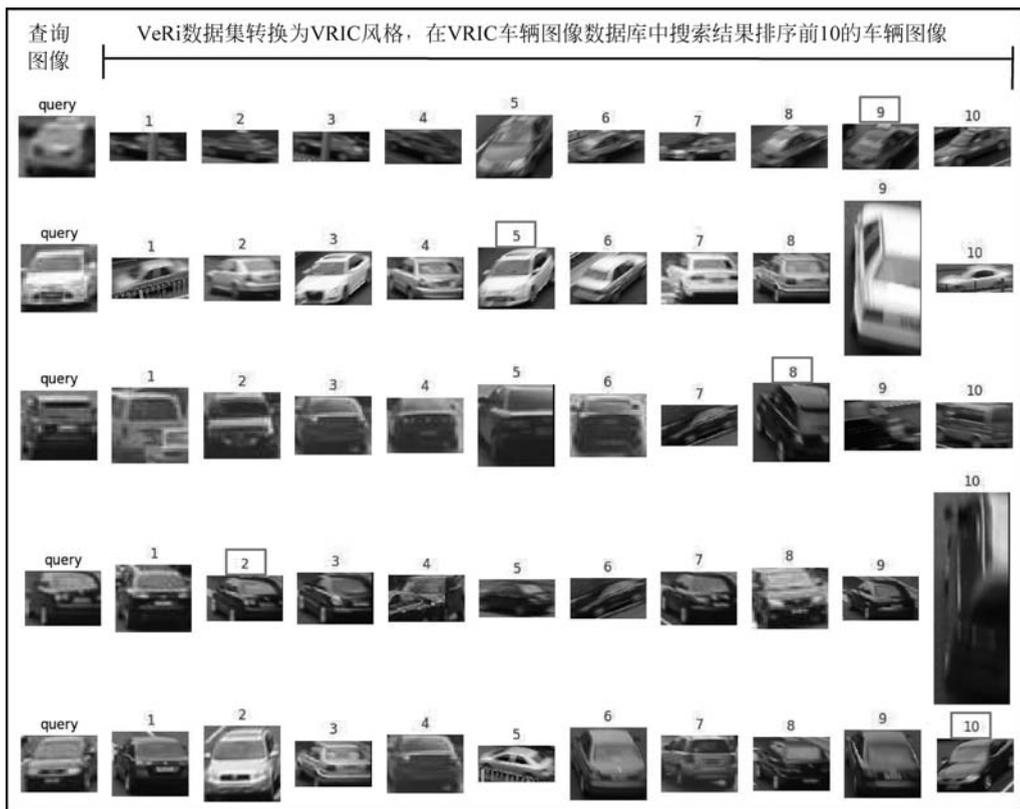


图 5-15 TLSA 在 VRIC 数据集上的车辆图像检索结果示例

3) 特征提取网络的影响

为了获得更好的车辆图像检索效果,本章在获取车辆特征时将 ResNet-50 替换为 CMNet-50,并引入二分支的局部最大池化(LMP)方法,形成最终的特征提取网络 CMNet-50-LMP。图 5-16 分别在 VeRi 数据集和 VRIC 数据集上对比了三种特征提取网络(ResNet-50、CMNet-50 和 CMNet-50-LMP)在车辆图像检索任务中的性能表现(mAP 精度)。对于 VeRi 数据集,从图 5-16(a)中可以看到,无论是直接转换的方法(Direct Transfer),

还是基于车辆图像迁移的方法(TLSA),当使用 CMNet-50-LMP 提取车辆特征时,车辆图像检索的效果都是最佳的。对于 VRIC 数据集,从图 5-16(b)中可以得到同样的结论。

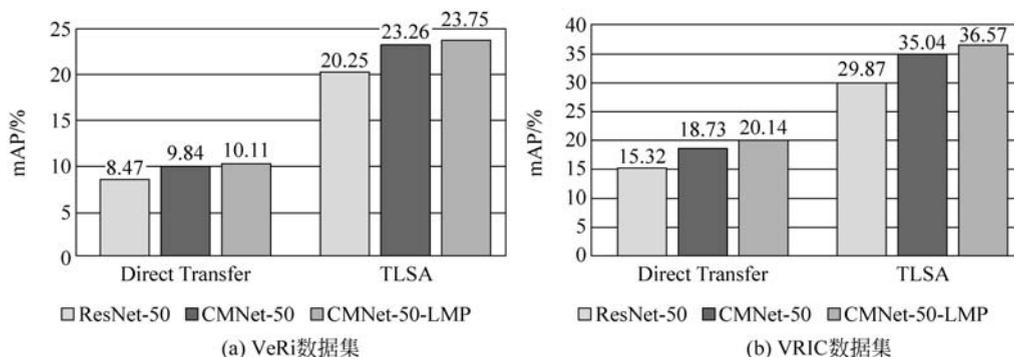


图 5-16 不同特征提取方法下域适应的表现

另外,将表 5-4 中的 Baseline 作为本章及后续章节中用于对比车辆图像检索任务的基准方法,该 Baseline 是基于身份学习的车辆图像检索方法。

5.5.5 车辆图像检索方法对比

为了进一步评估车辆图像风格迁移后用于图像检索的效果,将本章提出的 TLSA 方法与目前最先进的行人图像检索和车辆图像检索方法进行比较。

1. 在 VeRi 数据集上的方法对比

本节在 VeRi 数据集上评估了 TLSA 与其他 16 个行人或车辆图像检索模型的性能表现。其中,LOMO(Local Maximal Occurrence,局部最大发生)是一种利用手工绘制的局部特征解决行人检索的方法,旨在解决视角与光照变化引起识别效率不高的问题。BOW-SIFT 与 BOW-CN 都是基于视觉词汇向量 BoW(Bag-of-Words,词袋)模型搜索图像的方法,其中 BOW-SIFT 提出了一种贝叶斯合并的方法,通过降低词汇相关性的影响提高检索准确率,BOW-CN 将无监督的 BoW 描述子用于行人图像检索并获得了有竞争力的准确率。对于深度学习的方法,DGD(Domain Guided Dropout,域引导随机失活)利用卷积神经网络学习特定域的特征,FACT(Fusion of Attributes and Color Features,属性与颜色特征融合)与 NuFACT(Null-space based Fusion of Color and Attribute Feature,基于零空间的颜色与属性特征融合)分别将车辆的颜色、纹理与高级语义信息相结合,FACT+Plate-SNN+STR(Spatiotemporal Relations,时空关系)与 PROVID(Progressive Vehicle Re-Identification,渐进式车辆重新识别)将 FACT 与车牌验证相结合形成由粗到精、由近到远的渐进式车辆搜索方法,这些方法在一定程度上均获得了显著的效果。CVGAN(Cross-View Generative Adversarial Network,多视图生成对抗网络)与 SCCN-Ft + CLBL-8-Ft(Spatially Concatenated ConvNet+CNN-LSTM Bi-directional Loop,空间级联卷积网络+长短时记

忆双向循环网络)分别利用不同的生成对抗网络根据单视角的输入图像生成不同视角的车辆图像,ABLN-Ft-16(Adversarial Bi-directional LSTM Network,双向对抗长短时记忆网络)则提出利用双向对抗长短时记忆网络推断车辆所有视角的全局特征,VAMI(Viewpoint-aware Attentive Multi-view Inference,基于单视角注意力的多视图推理)使用单视角特征和注意力映射生成多视角特征,VAMI+STR则引入时空信息提升图像生成的质量,这些方法都是通过数据增强策略提升车辆的检索性能。JFSDL(Joint Feature and Similarity Deep Learning,联合特征与相似度的深度学习)应用孪生网络同时提取输入车辆图像对的特征,GSTE loss W/mean VGGM(Group-Sensitive-Triplet Embedding,组内敏感的三元组嵌入)提出使用三联网络学习车辆的细粒度特征,这两个方法通过提升特征质量提高了车辆检索效果。Siamese-CNN+Path-LSTM提出了一个两级框架,将视觉时空路径信息融合到车辆图像检索结果中,并验证了该方法的有效性。

在VeRi数据集上测试时,各种模型使用的训练集的情况如表5-3左侧所示。表5-5显示了TLSA方法在VeRi数据集上与其他方法的对比情况。其中,Baseline为非跨域场景下的车辆图像检索方法,是本章及后续章节对比的基准方法,TLSA为本章提出的基于迁移学习的车辆图像检索方法,其他方法为目前最先进的行人图像检索和车辆图像检索方法。从表中可以看到,TLSA获得了23.75%的mAP,同时Rank@1、Rank@5和Rank@20精度分别为44.26%、52.60%和75.86%。

表 5-5 TLSA 与经典的车辆图像检索方法在 VeRi 数据集上的结果比较(%)

| 方 法 | mAP | Rank@1 | Rank@5 | Rank@20 |
|-----------------------|-------|--------|--------|---------|
| BOW-SIFT | 1.51 | 1.91 | 4.53 | — |
| LOMO | 9.03 | 23.89 | 40.32 | 58.61 |
| BOW-CN | 12.20 | 33.91 | 53.69 | — |
| DGD | 17.92 | 50.70 | 67.52 | 79.93 |
| FACT | 18.54 | 52.35 | 67.16 | 79.97 |
| XVGAN | 24.65 | 60.20 | 77.03 | 88.14 |
| ABLN-Ft-16 | 24.92 | 60.49 | 77.33 | 88.27 |
| SCCN-Ft+CLBL-8-Ft | 25.12 | 60.83 | 78.55 | 89.79 |
| FACT+Plate-SNN+STR | 27.77 | 61.44 | 78.78 | — |
| NuFACT | 48.47 | 76.76 | 91.42 | — |
| VAMI | 50.13 | 77.03 | 90.82 | 97.16 |
| PROVID | 53.42 | 81.56 | 95.11 | — |
| JFSDL | 53.53 | 82.90 | 91.60 | — |
| Siamese-CNN+Path-LSTM | 58.27 | 83.49 | 90.04 | — |
| GSTE loss W/mean VGGM | 59.47 | 96.24 | 98.97 | — |
| VAMI+STR | 61.32 | 85.92 | 91.84 | 97.70 |
| Baseline | 60.01 | 89.63 | 95.47 | 96.84 |
| TLSA | 23.75 | 44.26 | 52.60 | 75.86 |

与 Baseline 方法相比,TLSA 在 VeRi 数据集上的表现大幅度下降,这与预期是相符的,这是由不同数据域中车辆图像风格的差异导致的,即使通过迁移学习将源域中的图像转换为目标域中图像的风格,但仍然具有很大的差距。

相对于非深度学习的方法,TLSA 的 mAP 比 BOW-SIFT、LOMO 和 BOW-CN 分别高 22.24%、14.72% 和 11.55%,说明图像风格迁移方法用于车辆的检索任务是可行的。

与其他基于深度学习的方法(如 DGD、FACT、XVGAN、ABLN-Ft-16、SCCN-Ft + CLBL-8-Ft、FACT+Plate-SNN+STR、NuFACT、VAMI、PROVID、JFSDL、Siamese-CNN + Path-LSTM、GSTE loss W/mean VGGM 和 VAMI+STR)相比,在 mAP 精度上,TLSA 方法的表现优于 DGD 和 FACT,却低于其他所有方法,这说明基于图像风格迁移的 TLSA 方法是有效的,同时可以看到,跨域场景下的车辆图像检索仍然是一个具有挑战性的任务。

2. 在 VRIC 数据集上的方法对比

在 VRIC 数据集上测试时,各种模型使用的训练集的情况如表 5-3 右侧所示。表 5-6 显示了 TLSA 方法在 VRIC 数据集上与 Baseline 方法的对比情况。从比较结果看到,TLSA 方法在 VRIC 数据集上的表现同样大幅度下降,但是仍然获得了 32.57% 的 mAP,并且 Rank@1、Rank@5 和 Rank@20 精度分别为 34.18%、48.69% 和 59.53%,表明了基于图像风格迁移的车辆图像检索方法(TLSA)是有效的,但仍然需要进一步深入研究。

表 5-6 TLSA 在 VRIC 数据集上的检索结果(%)

| 方 法 | mAP | Rank@1 | Rank@5 | Rank@20 |
|----------|-------|--------|--------|---------|
| Baseline | 64.13 | 60.05 | 83.26 | 90.62 |
| TLSA | 32.57 | 34.18 | 48.69 | 59.53 |

5.6 本章小结

本章研究了车辆图像检索技术,针对跨域场景下图像风格变化显著等问题,提出了一种基于迁移学习场景自适应的车辆图像检索方法。首先介绍了车辆迁移生成对抗网络的算法原理和训练方法;然后研究了基于图像风格迁移的车辆图像检索算法,在特征学习过程中重点阐述局部最大池化方法;最后在 VeRi 数据集和 VRIC 数据集上分别对车辆图像风格迁移的效果和车辆图像检索的性能进行评估,验证了本章方法的有效性。

参考文献

- [1] LIU X C, LIU W, MA H D, et al. Large-scale Vehicle Re-identification in Urban Surveillance Videos [C]. Proceedings of the 2016 IEEE International Conference on Multimedia and Expo (ICME), 2016: 1-6.

- [2] LIU X C, LIU W, MEI T, et al. A Deep Learning-based Approach to Progressive Vehicle Re-identification for Urban Surveillance[C]. Proceedings of the 2016 European Conference on Computer Vision(ECCV), 2016: 869-884.
- [3] HADSELL R, CHOPRA S, LECUN Y. Dimensionality Reduction by Learning an Invariant Mapping [C]. Proceedings of the 2016 IEEE Computer Society Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 1735-1742.
- [4] BROMLEY J, GUYON I, LECUN Y, et al. Signature Verification Using a "Siamese" Time Delay Neural Network[J]. International Journal of Pattern Recognition and Artificial Intelligence, 1993, 7 (4): 669-688.
- [5] PASZKE A, GROSS S, CHINTALA S, et al. Automatic Differentiation in PyTorch[C]. Proceedings of the 2017 Neural Information Processing Systems(NIPS), 2017.
- [6] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [7] LIAO S C, HU Y, ZHU X Y, 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(CVPR), 2015: 2197-2206.
- [8] ZHENG L, WANG S J, ZHOU W G, et al. Bayes Merging of Multiple Vocabularies for Scalable Image Retrieval [C]. Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2014: 1963-1970.
- [9] ZHENG L, SHEN L Y, TIAN L, et al. Scalable Person Re-identification: A Benchmark [C]. Proceedings of the 2015 IEEE International Conference on Computer Vision(ICCV), 2015: 1116-1124.
- [10] XIAO T, LI H S, QUYANG W L, et al. Learning Deep Feature Representations with Domain Guided Dropout for Person Re-identification [C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 1249-1258.
- [11] ZHOU Y, SHAO L. Cross-view GAN Based vehicle Generation for Re-identification [C]. Proceedings of the 2017 British Machine Vision Conference(BMVC), 2017.
- [12] ZHOU Y, SHAO L. Vehicle Re-identification by Adversarial Bi-directional LSTM Network [C]. Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision(WACV), 2018: 653-662.