

# 基于多源域迁移学习的行人重识别探讨

张孟思

(湖南涉外经济学院 长沙 410205)

**摘要** 文中以多源域迁移学习方法作为支持,对行人重识别策略的应用进行了分析,包括多源域迁移学习算法、行人重识别现状及其发展需求、多源域迁移学习方法下的行人重识别策略。以期多源域迁移学习方法的合理应用与行人重识别质量的提升提供科学参考,对提高监管区域监控效能具有积极意义。

**关键词:** 多源域迁移学习;行人重识别;监控效能

**中图分类号** TP311

## Exploration of Person Re-Identification Based on Multi-source Domain Transfer Learning

ZHANG Mengsi

(Hunan International Economics University, Changsha 410205, China)

**Abstract** This paper analyzes the application of person re-identification strategy based on multi-source domain transfer learning. This includes multi-source domain transfer learning algorithms, the current status and development needs of unsupervised person re-identification, and unsupervised person re-identification strategies under multi-source domain transfer learning methods. I hope that through this analysis, scientific references can be provided for the rational application of multi-source domain transfer learning methods and the improvement of unsupervised person re-identification quality. It has positive significance to improve the monitoring efficiency of the supervision area.

**Key words** Multi-source domain transfer learning, Person re-identification, Monitoring efficiency

## 0 引言

在以多源域迁移学习方法为基础实施行人识别的过程中,研究者与技术人员首先需要科学掌握多源域学习方法,包括其基本情况与主要优势,还应进一步了解行人重识别的现状及其发展需求。再以此为依据,将多源域迁移学习方法合理应用到行人重识别中,对目标函数进行合理的设计、简化与优化,并通过投影矩阵和距离度量的方式来实现行人重识别。这样才可以科学处理监控区域摄像头拍摄到的行人数据,以满足监管区域实际的监控需求。

## 1 多源域迁移学习方法

### 1.1 基本情况

多源域迁移学习方法是基于传统迁移学习方法发展而来的一种新型迁移学习方法。该方法所涉及的源域有很多,将这些源域合理迁移到相应的知识学习中,可对目标域学习起到辅助作用。通过多个源域信息的合理利用与融合,可进一步提升迁移学习的效率及其质量,满足现代深度学习模型的实际应用及其发展需求<sup>[1]</sup>。尤其是在目标域的

数据比较有限或源域差异性较大的情况下,该方法的应用优势十分突出。

### 1.2 主要优势

目前,从深度学习领域来看,多源域迁移学习方法的主要优势表现在以下几方面。(1)通过合理利用各相关领域知识,可使迁移学习目标领域中的学习任务具有更高的基线性能,以此来为其学习模型提供更高起点。(2)通过多源域知识的合理应用,该方法可缩短模型学习时间,提高开发效率。(3)该学习方法可在相似任务或领域之间达到良好的迁移信息共享效果,进一步降低模型的学习成本。综上,多源域迁移学习方法在现代的深度学习模型中得到了越来越广泛的应用,且在很多应用领域都发挥了显著优势。

## 2 行人重识别现状及其发展需求

### 2.1 行人重识别现状

在视频监控领域,不同摄像头摄取的环境信息通常存在一定差异,如拍摄距离、光照强度、拍摄角度等,都会使其摄取的信息各不相同。在这样的情况下,通过不同摄像头

**基金项目:**湖南省教育厅资助科研项目:基于遮挡下特征融合的行人重识别算法研究(22C066)

**作者简介:**张孟思(1996—),硕士,讲师,研究方向为计算机技术。

摄取到的行人视频也会呈现出明显差异。为了让同一个行人存在明显差异的视频得到良好匹配,技术人员通常需要采取人工的方式来标记训练数据,在两个摄像头之间建立联系。目前,以度量学习模型为支持的视频行人重识别算法在很多被标记的行人视频处理中得到了良好应用,且表现出了较好的鉴别能力。

## 2.2 行人重识别发展需求

度量学习模型在视频行人重识别中发挥出了较好的应用效果,但在真实视频监控环境里,此种方法的实现却具有难度。因为真实视频监控网络里的摄像机数量较多,且随着应用场景的扩大与应用需求的提升,设置在其中的摄像机数量也将不断增加。在这样的情况下,如果再想从任意两个不具有重叠关系的视频里找出共同出现的大量行人,并对其进行标记,则是一件费时费力的事情。在某些情况下,两个摄像机之间根本就不会有大量行人共同出现。基于此,度量学习模型在视频监控领域中发挥的作用也越来越有限。为解决这一问题,实现视频监控领域行人重识别技术的良好发展,需要合理引入多源域迁移学习支持下的学习技术,使学习模型的训练不再依赖于目标数据标签<sup>[2]</sup>。基于此,技术人员可将度量学习模式下的行人重识别技术作为基础,将多源域迁移学习方法合理融入其中,使模型在训练时对目标数据集标记达到“零需求”。这样才可以满足现代视频监控领域对行人重识别技术的实际应用需求,促进此类技术、算法和整个视频监控领域的良好发展。

## 3 多源域迁移学习方法下的行人重识别策略

### 3.1 目标函数设计

在多源域迁移学习方法的支持下,本文提出了一种形式的度量学习模型,其旨在将被标记的行人视频训练获得的鉴别力较强的距离度量向没有标记形式的目标行人监控视频数据中迁移。为便于后续讨论,研究者首先需对一些常用术语进行定义。(1)源数据集(i),即公开发布的带有标记的行人监控视频数据集。(2)目标数据集(ii),即用来测试模型的数据集。在具体应用时,主要目的是将在源数据集中学习的距离度量有效迁移并应用在目标数据集中。在该模型设计中,应注意以下两个问题。(1)怎样通过(i)将源数据集与目标数据集的差异性对迁移结果的影响削弱。(2)怎样确保(ii)训练获得的模型鉴别力足够强。为解决这两个问题,可将源数据集与目标数据集中的数据分布拉近<sup>[3]</sup>。如果负样本难以区分的数据集鉴别信息更多,研究者则可合理利用具有更多鉴别信息的数据集,这将有利于行人重识别形式的距离度量学习。图1为多源域迁移学习方法支持下的度量学习模型主要构建思想。

假设具有标记形式的源数据集为  $X = [x_1, \dots, x_i, \dots, x_{N_1}]$ , 其中  $x_i$  代表第  $i$  个行人在监控视频中的特征向量,  $N_1$  代表  $X$  里的样本总数。假设没有标记形式的目标数据集是  $Y = [y_1, \dots, y_j, \dots, y_{N_2}]$ , 其中  $y_j$  代表第  $j$  个行人在监控视频里的特征

向量;  $N_2$  代表  $Y$  里的样本总数。假设源数据集与目标数据集差异削弱投影矩阵是  $W$ , 特征学习中的距离度量矩阵是  $V$ 。则可计算多源域迁移学习方法下的度量模型目标函数, 如式(1)所示:

$$\min_{W,V} E_{trans}(X,Y,W) + \alpha E_{pos}(X,W,V) + \beta E_{disc}(X,W,V) + \gamma E_{reg}(W) \quad (1)$$

$$s.t. \quad \|v_i\|_2 \leq 1$$

其中,  $\alpha, \beta, \gamma$  均为平衡参数, 其主要作用是实现各个功能项的有效平衡。  $\|v_i\|_2 \leq 1$  属于约束条件, 其主要作用是对  $V$  量级加以限制。

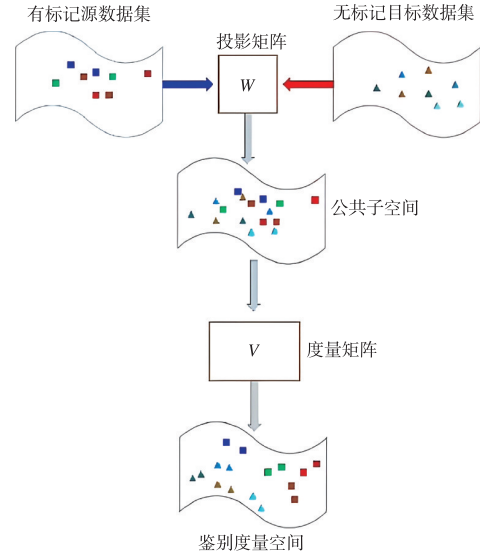


图1 多源域迁移学习方法支持下的度量学习模型主要构建思想

$E_{trans}(X,Y,W)$  代表削弱源数据集与目标数据集两者差异性的功能项, 其主要作用是让源数据集以及目标数据集中的数据可以尽可能相似地分布在公共子空间里<sup>[4]</sup>。其定义式如式(2)所示:

$$E_{trans}(X,Y,W) = \left\| \frac{1}{N_1} \sum_{i=1}^{N_1} W^T x_i - \frac{1}{N_2} \sum_{j=1}^{N_2} W^T y_j \right\|_2^2 \quad (2)$$

$E_{pas}(X,Y,W)$  代表收紧源数据集以及正确匹配样本之间的功能项, 其主要作用是使该模型学习到的距离度量在公共子空间里能透射到的正确匹配视频距离缩小。其定义式如式(3)所示:

$$E_{pas}(X,W,V) = \frac{1}{|S|} \sum_{\langle i,j \rangle \in S} \left\| V^T (W^T x_i - W^T y_j) \right\|_2^2 \quad (3)$$

其中,  $S$  代表带有标记形式的源数据集里正确匹配视频对所组成的全部集合,  $|S|$  代表一个集合里的元素总数。

$E_{disc}(X,Y,W)$  代表鉴别具有标记形式的源数据集功能项, 其主要作用是使该模型学习到的距离度量能令公共子空间里正确匹配行人视频具有较错误匹配的距离小。其定义式如式(4)所示:

$$E_{disc}(X,W,V) = \frac{1}{|G|} \sum_{\langle i,j,k \rangle \in G} \left\| V^T (W^T x_i - W^T x_j) \right\|_2^2 - \left\| V^T (W^T x_i - W^T x_k) \right\|_2^2 \quad (4)$$

其中,  $G$  代表难分负样本组成的难分三元组集合, 其构造规则是用  $x_i, x_j, x_k \in R$  分别代表第  $i$  个行人、第  $j$  个行人以及第  $k$  个行人在监控视频片段里的特征向量, 用  $y_i, y_j, y_k$  分别代表上述三个视频片段所对应的标签, 且有  $y_i = y_j, y_i \neq y_k$ 。若  $\|x_i - x_k\|_2 < \|x_i - x_j\|_2 + 1$ , 则将  $x_k$  叫做  $x_i$  目标样本中的难分负样本, 将  $\langle x_i, x_j, x_k \rangle$  叫做难分三元组。

$E_{reg}(X, Y, W)$  为正则化功能项, 其主要作用是对  $W$  参数实施正则化处理。其定义式如式(5)所示:

$$E_{reg}(W) = \|W\|_F^2 \quad (5)$$

此次构建的模型主要对一个投影矩阵以及一个距离度量进行联合学习, 其中投影矩阵的主要作用是将源数据集以及目标数据集分布差异消除; 距离度量的主要作用是使模型具有更高鉴别能力<sup>[5]</sup>。在该模型的学习过程中, 不需要应用目标数据集中具有标记形式的数据, 通过源数据集了的数据对其进行训练即可, 因此其实际训练环境的扩展性更佳。

### 3.2 目标函数简化

为简化目标函数, 可先缩写其中的几项。

(1) 对式(2)进行简化, 如式(6)所示:

$$E_{trans}(X, Y, W) = \|W^T Z_1\|_2^2 \quad (6)$$

其中,  $Z_1 = \frac{1}{N_1} \sum_{i=1}^{N_1} x_i - \frac{1}{N_2} \sum_{j=1}^{N_2} y_j$ 。

(2) 对式(3)进行简化, 如式(7)所示:

$$E_{pas}(X, Y, W) = \frac{1}{|S|} \|V^T W^T Z_2\|_F^2 \quad (7)$$

其中,  $Z_2$  为中间矩阵, 每一列都是  $S$  里的一组正确匹配对之差, 即  $\langle x_i, x_j \rangle$  之差  $x_i - x_j$ 。

(3) 对式(4)进行简化, 如式(8)所示:

$$E_{disc}(X, W, V) = \frac{1}{|G|} \left( \|V^T W^T Z_3\|_F^2 - \|V^T W^T Z_4\|_F^2 \right) \quad (8)$$

其中,  $Z_3$  和  $Z_4$  均为中间矩阵。在  $Z_3$  中, 每一列都是  $G$  里的一组难分三元组对之差, 即  $\langle x_i, x_j \rangle$  之差  $x_i - x_j$ ; 在  $Z_4$  中, 每一列都是  $G$  里的一组难分三元组对之差, 即  $\langle x_i, x_j, x_k \rangle$  之差  $x_i - x_k$ 。

将式(6)、式(7)和式(8)导入到式(1)中, 可对目标函数进行重写, 如式(9)所示:

$$\begin{aligned} \min_{W, V} & \|W^T Z_1\|_2^2 + \alpha \frac{1}{|S|} \|V^T W^T Z_2\|_F^2 + \beta \frac{1}{|G|} \\ & \left( \|V^T W^T Z_3\|_F^2 - \|V^T W^T Z_4\|_F^2 \right) + \gamma \|W\|_F^2 \\ \text{s.t.} & \|v_i\|_2 \leq 1 \end{aligned} \quad (9)$$

进一步合并式(9)里的相关项, 如式(10)所示:

$$\begin{aligned} \min_{W, V} & \|W^T Z_1\|_2^2 + \alpha \frac{1}{|S|} \|V^T W^T Z_5\|_F^2 + \beta \frac{1}{|G|} \|V^T W^T Z_4\|_F^2 + \gamma \|W\|_F^2 \\ \text{s.t.} & \|v_i\|_2 \leq 1 \end{aligned} \quad (10)$$

其中,  $Z_5 = [\sqrt{\alpha/|S|} Z_2, \sqrt{\beta/|G|} Z_3]$ 。

目标函数在式(10)里属于非凸函数, 不能直接优化  $W$  以及  $V$ 。在对多个变量进行优化的过程中, 研究者可引入

变量来松弛处理目标函数, 这是一项非常有效的措施。基于此, 研究者引入  $A$  和  $B$  两个变量, 对式(10)进行了松弛处理, 如式(11)所示:

$$\begin{aligned} \min_{W, V} & \|W^T Z_1\|_2^2 + \|V^T A\|_F^2 - \beta \frac{1}{|G|} \|V^T B\|_F^2 + \gamma \|W\|_F^2 + \\ & \tau \left( \|W^T Z_5 - A\|_F^2 + \|W^T Z_4 - B\|_F^2 \right) \\ \text{s.t.} & \|v_i\|_2 \leq 1 \end{aligned} \quad (11)$$

其中,  $\tau$  为常数。对于式(11)里的目标函数, 在具体计算时, 可采用  $W$  以及  $V$  迭代更新的方式对其实施进一步的优化处理。

### 3.3 目标函数优化

根据上述的目标函数处理方案, 在对本文所提出模型中的目标函数进行优化时, 研究者采取了交替迭代优化法, 通过对多个变量进行更新来优化该模型的目标函数。在对其中的某个或某些变量进行优化时, 研究者可使其他变量均保持不变。在此过程中, 可以按照3个小问题对式(11)实施交替求解。

(1) 对  $A$  和  $B$  进行更新。使式(11)中的  $W$  和  $V$  保持不变, 仅对其中的  $A$  和  $B$  进行更新处理。在对  $A$  和  $B$  进行更新之前, 首先进行  $W$  以及  $V$  的初始化处理, 将  $W$  初始为单位矩阵, 对  $V$  实施初始化处理, 如式(12)所示:

$$\begin{aligned} \min_V & \|V^T W^T Z_5\|_F^2 - \beta \frac{1}{|G|} \|V^T W^T Z_4\|_F^2 \\ \text{s.t.} & V^T V = I \end{aligned} \quad (12)$$

这里的  $I$  为单位矩阵, 在具体处理时, 可构造一个拉格朗日函数, 将这个函数对于与  $V$  的偏导数设置为 0, 这样可获得式(13)所示:

$$\left( W^T Z_5 Z_5^T W - \beta \frac{1}{|G|} W^T Z_4 Z_4 W \right) V = \lambda V \quad (13)$$

通过特征值分解方法的合理应用, 可有效解决式(13)中的问题, 即将  $V$  初始化为与最小特征值中的前  $k$  个相对应的特征向量组成的矩阵。令  $W$  以及  $V$  固定不变, 可用式子表示式(11)目标函数和  $A$  或  $B$  之间的相关部分, 如式(14)、式(15)所示:

$$\min_A \|V^T A\|_F^2 + \tau \|W^T Z_5 - A\|_F^2 \quad (14)$$

$$\min_B -\beta \frac{1}{|G|} \|V^T B\|_F^2 + \tau \|W^T Z_4 - B\|_F^2 \quad (15)$$

先更新  $A$ , 求式(14)里  $A$  的偏导, 并将其设置为 0, 由此可得式(16):

$$A = \tau (V V^T + \tau I)^{-1} W^T Z_5 \quad (16)$$

通过上述方法对式(15)进行求解, 由此可得出更新之后的  $B$ , 如式(17)所示:

$$B = \tau \left( \tau I - \frac{\beta}{|G|} V V^T \right)^{-1} W^T Z_4 \quad (17)$$

(2) 对  $V$  进行更新。使式中的  $A, B$  以及  $W$  保持不变, 仅对其中的  $W$  进行更新。将式(11)目标函数以及  $V$  有关的部分如式(18)所示:

$$\begin{aligned} \min_V & \|V^T A\|_F^2 - \beta \frac{1}{|G|} \|V^T B\|_F^2 \\ \text{s.t.} & \|v_i\|_2 \leq 1 \end{aligned} \quad (18)$$

为解决式(18)的问题,研究者引入另一变量  $U$ ,并改写式(18),如式(19)所示:

$$\begin{aligned} \min_V & \|V^T A\|_F^2 - \beta \frac{1}{|G|} \|V^T B\|_F^2 \\ \text{s.t.} & V = U, \|U_i\|_2 \leq 1 \end{aligned} \quad (19)$$

对于式(19),研究者主要通过 ADMM 算法来求解,如式(20)所示:

$$\begin{cases} V = \arg \min_V \|V^T A\|_F^2 - \beta \frac{1}{|G|} \|V^T B\|_F^2 + \rho \|V - U + T_1\|_F^2 \\ U = \arg \min_U \rho \|V - U + T_1\|_F^2, \text{s.t.} \|U_i\|_2 \leq 1 \\ T_1 = T_1 + V - U \end{cases} \quad (20)$$

这里的  $T_1$  初始值是零矩阵。3 个步骤交换更新,一直到收敛为止。

(3)对  $W$  进行更新。使式中的  $A$ ,  $B$  以及  $V$  保持不变,式(11)目标函数和  $W$  之间的相关部分按以下方法进行更新。将带有标记形式的  $X$  源数据集以及没有标记形式的  $Y$  目标数据集输入该模型;将  $W$  投影矩阵以及  $V$  距离变量输出;将  $W$  初始化为单位阵,通过式(12)对  $V$  进行初始化处理(首先将  $W$  以及  $V$  固定,通过式(16)、式(17)对  $A$  和  $B$  进行更新;将  $A$ ,  $B$  以及  $W$  固定,通过式(20)对  $V$  进行更新;将  $A$ ,  $B$  以及  $V$  固定,对  $W$  进行更新,如式(21)所示):

$$\min_W \|W^T Z_1\|_2^2 + \tau (\|W^T Z_5 - A\|_F^2 + \|W^T Z_4 - B\|_F^2) + \gamma \|W\|_F^2 \quad (21)$$

求式(21)中  $W$  的偏导,再将其设置为 0,可得出结果如式(22)所示:

$$W = \tau (Z_1 Z_1^T + \tau Z_5 Z_5^T + \tau Z_4 Z_4^T + \gamma I) - 1 + \gamma \|W\|_F^2 \quad (22)$$

通过这样的方式,可合理优化该模型算法,以充分发挥多源域迁移学习法的应用优势,满足视频监控领域中的行人重识别处理需求。

### 3.2 基于投影矩阵与距离度量的行人重识别

在通过投影矩阵以及距离度量对视频监控中的行人视频进行重识别处理时,研究者可将目标数据集里测试到的一段行人视频设置为  $y_p$ ,将目标数据集里候选的行人监控视频集合设置为  $Y_c = \{y_g | g = 1, 2, \dots, q\}$ ,  $q$  代表候选集里的行人监控视频总数。通过上述训练获得的  $W$  投影矩阵以及  $V$  距离度量矩阵,便可对  $y_p$  和  $Y_c$  做出合理匹配。以下是具体的匹配步骤。(1)利用上述训练获得的  $W$  投影矩阵在公共子空间里完成目标数据集里的测试以及候选监控视频投影,

投影之后,前者的样本用  $y_p' = W^T y_p$  表示;后者的样本用  $y_g' = W^T y_g$  表示。(2)对投影之后测试和全部候选人监控视频间距  $d = \{d_1, d_2, \dots, d_q\}$  进行计算,如式(23)所示:

$$d(y_p', y_g') = \|V^T (y_p' - y_g')\|_2^2 \quad (23)$$

按升序排列  $d = \{d_1, d_2, \dots, d_q\}$ ,将最小距离的候选监控视频作为测试视频中的正确匹配<sup>[6]</sup>。通过这样的方式,可对监控视频中的行人做出有效的重识别处理,以充分满足特定行人的监控需求。图 2 为投影矩阵以及距离度量矩阵条件下的行人监控视频匹配示意图。

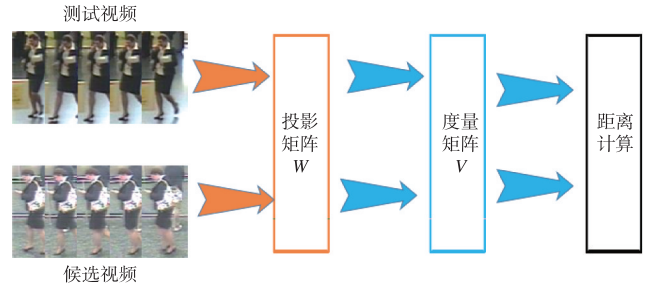


图 2 投影矩阵以及距离度量矩阵条件下的行人监控视频匹配示意图

## 4 结语

综上所述,在对监控视频录像中的行人进行重识别处理时,传统的迁移学习方法不能对多源域形式的视频做出有效处理。为解决这一问题,研究者与技术人员需要引入更加先进的多源域迁移学习方法,借助投影矩阵以及距离度量矩阵对获取到的监控视频进行处理,以实现目标行人的重识别。这将有利于现代监控视频处理质量的提升及其监控作用的发挥,从而最大限度地满足监控区域的实际监督与管理需求,达到良好的行人重识别效果。此外,这可以对现代深度学习模型的应用与视频监控效果的提升起到有力的技术支持。

### 参考文献

- [1] 周靖宇. 监控视频中的跨模态行人重识别研究[D]. 陕西: 西安电子科技大学, 2022.
- [2] 胡彬, 潘雨, 丁卫平, 等. 基于迁移学习的行人再识别[J]. 图学学报, 2018(5): 886-891.
- [3] 范星. 智能视频监控中的行人重识别方法研究[D]. 浙江: 浙江大学, 2020.
- [4] 黄杨茹. 基于迁移学习的行人重识别算法研究[D]. 北京: 北京交通大学, 2020.
- [5] 唐鼎. 基于度量学习和迁移学习的视频行人重识别研究[D]. 河南: 河南大学, 2020.
- [6] 邱耀儒. 基于迁移学习的行人重识别研究[D]. 广东: 广东工业大学, 2020.