第 12 卷 第 1 期
2023 年 1 月
集 成 技 术
JOURNAL OF INTEGRATION TECHNOLOGY
Vol. 12 No. 1
Jan. 2023

引文格式：
朱利, 林欣, 徐亦飞, 等. 基于城市信息单元和差异注意力的多层行人重识别技术 [J]. 集成技术, 2023, 12(1): 91-104.
Zhu L, Lin X, Xu YF, et al. Multi-level person re-identification based on urban information unit and diff attention scheme [J]. Journal of Integration Technology, 2023, 12(1): 91-104.

基于城市信息单元和差异注意力的多层行人重识别技术

朱 利[1*] 林 欣[1] 徐亦飞[1] 刘 真[2] 马 英[3]

[1]（西安交通大学电信学部软件学院 西安 710049）
[2]（北京交通大学计算机科学与信息学院 北京 100091）
[3]（国家信息中心 北京 100038）

摘 要 在现实的智慧城市安全场景中，传统的行人重识别方法已经难以满足复杂多样的识别任务要求。为实现多层次的行人重识别，该文提出将行人重识别技术与多层次的城市信息单元深度融合。在行人重识别任务中，现有的模型和注意力只关注鲁棒特征的学习，而该文基于特征向量差异，提出了差异注意力模块，以增强深度特征的判别力。结合差异注意力模块，该文开发了与多种骨干模型适配的差异注意力框架。此外，该文还提出了联合训练和单独训练两种训练策略。与其他行人重识别方法相比，差异注意力框架和训练策略在 Market-1501、CUHK03 和 MSMT17 数据集上均取得了更优的性能。

关键词 行人重识别；城市信息单元；差异注意力；距离函数；深度学习
中图分类号 TP 391.41 **文献标志码** A doi: 10.12146/j.issn.2095-3135.20220712001

Multi-level Person Re-identification based on Urban Information Unit and Diff Attention Scheme

ZHU Li[1*] LIN Xin[1] XU Yifei[1] LIU Zhen[2] MA Ying[3]

[1](*School of Software Engineering*, *Xi'an Jiaotong University*, *Xi'an* 710049, *China*)
[2](*School of Computer and Information Technology*, *Beijing Jiaotong University*, *Beijing* 100091, *China*)
[3](*State Information Center*, *Beijing* 100038, *China*)
Corresponding Author: *zhuli@xjtu.edu.cn*

Abstract The traditional person re-identification methods are difficult to independently cope with the complex and diverse recognition tasks in the security scenario of smart city in practice. In order to meet the

收稿日期：2022-07-12 **修回日期**：2022-08-28
基金项目：国家重点科研项目（2019YFB2102500）
作者简介：朱利（通讯作者），教授，研究方向为机器学习与计算机网络，E-mail：zhuli@xjtu.edu.cn；林欣，硕士研究生，研究方向为计算机视觉与行人重识别；徐亦飞，副教授，研究方向为视频理解与图像处理技术；刘真，副教授，研究方向为社交网络与社会计算、推荐系统、知识图谱、并行与分布式计算；马英，高级工程师，研究方向为政务信息管理。

needs of multi-level person re-identification, the deep integration of person re-identification and multi-level urban information units is proposed. Existing models and attentions for person re-identification tasks only focus on learning the robust features while neglecting the difference between features of pairs. Diff attention module is proposed to guide the network to learn a more discriminative attention map based on the difference of feature vectors. Taking the diff attention module, diff attention framework which matches many backbone models is developed. Two training strategies: joint training and separate training are proposed. Compared with other person re-identification methods, these framework and strategies have achieved excellent performance on Market-1501, CUHK03, and MSMT17 datasets.

Keywords person re-identification; urban information unit; diff attention; distance function; deep learning
Funding This work is supported by National Key Research and Development Project of China (2019YFB2102500)

1 引 言

行人重识别(Person Re-identification，Re-ID)是一个特殊的人员检索问题，近年来受到了工业界和学术界的广泛关注。Person Re-ID 的目的是在不同的时间、摄像机或场景中匹配一个特定的人，称为“查询人”。由于从图像、视频和文本描述中提取有鉴别性特征的方式不同，Person Re-ID 十分具有挑战性。此外，不同视角、背景杂波、姿势多样性和遮挡的存在为 Person Re-ID 任务带来了变化和不确定性。

随着公众安全的迫切需求和城市中监控摄像机数量的不断增加，在复杂城市环境中，如何匹配识别特定人物给智慧城市带来了严峻的挑战。在研究与实验中，传统行人重识别数据集的样本数量有限、风格单一，且 Re-ID 任务只是查询图像在图库中进行相似匹配。而在现实的行人重识别任务中，通过多种渠道收集的行人图像数量庞大、风格迥异、相似匹配难度大。因此，单一的行人重识别技术难以应对复杂的识别需求。为提高行人重识别技术的实用性，本文提出将行人重识别技术与多级城市信息单元深度融合，形成相似的层次结构，可以将任务的数据规模控制在一定范围内。该融合便于构建解决实际问题的概念模型，可将复杂的现实识别需求分解为多级城市信息单元框架下的多个明确的行人重识别子问题，从而使行人重识别技术满足智慧城市场景下的多层次行人重识别任务需要。

近年来，大量研究集中于利用深度神经网络进行行人重识别，识别效果良好[1-3]。相关学者还针对其训练技巧和性能提升进行了研究[4]，尝试将行人重识别技术与注意力机制相结合，以增强深度特征的辨别性，并抑制无用特征[5-9]。大多数注意力由有限感受野的全连接层或卷积层进行学习，但它们仅使用了单个图像信息。

现有的深度特征学习模型和注意力机制只关注深度特征与其对应样本数据之间的关系，而忽略了不同特征对之间的差异。实际上，通过深度特征学习方法解决行人重识别问题的核心是将检索问题转化为深度特征的相似匹配任务。然而，目前基于距离函数的深层网络一般都局限于特定的数据集或特定的识别任务。

本文设计了差异注意力模块解决特征相似性匹配任务，实现了基于深度特征向量对差异的注意力机制。为使差异注意力模块能够匹配多样的深度特征模型，且保证提取特征的多样性，本文提出了差异注意力框架。此外，还设计了两种不同的训练策略用于训练差异注意力模块和整个框架。

本文主要工作如下：

(1)将行人重识别技术与多级城市信息单元深度融合，形成相似的层次结构，使行人重识别技术能够满足智慧城市场景下的多层次行人重识别任务需求。

(2)指出基于深度特征表示的行人重识别问题的核心是特征向量之间的差异，提出差异注意力的思想，通过差异注意力选择更具有辨别力的特征。

(3)设计了差异注意力模块，用于实现基于深度特征差异的差异注意力机制。设计了差异注意力框架和两种不同的训练策略(联合训练和单独训练)以匹配不同的深度模型并对其进行训练。在 Market-1501、CUHK03 和 MSMT17 等行人重识别数据集上，与其他行人重识别特征表示方法相比，差异注意力的效果更好。

2　相关工作

在计算机视觉中，行人重识别是一项具有挑战性且十分复杂的任务。本节将讨论城市信息单元、与行人重识别相关的特征表示学习和面向有监督的行人重识别的注意力机制。

2.1　城市信息单元

根据城市行政区划，城市信息单元[10]在地理上分为网格、区域、街道和市辖区。每个城市信息单元包含基本的政府数据和社会传感器数据。其中，政府数据包括人口普查结果、社会经济指标、地图、街道等信息；社会传感器数据包括天气、温度、水质、交通流量、人流等信息。

一座城市包含一个或多个市政区，每个市政区包含一条或多条街道，街道又包含社区、小学、购物中心、公园等区域。根据纬度和经度，城市在地理上可被划分为多个网格。因此，城市信息单元有类似的层次结构：每个市政区级城市信息单元包含一个或多个街道级城市信息单元，每个街道级城市信息单元包含一个或多个区域级城市信息单元，每个区域级城市信息单元包含一个或多个网格级城市信息单元，网格级城市信息单元是最基础的城市信息单元层级。

2.2　特征表示学习

特征表示学习是从具有良好识别能力的行人重识别数据集中提取样本图像的特征向量。目前，主要有 4 种特征学习策略：全局特征、局部特征、辅助特征和视频特征[11-12]。其中，全局特征是从每个人物图像中提取全局的特征表示向量[1]；局部特征聚合了不同的零件级局部特征，便于为每个人物图像组合出一个新的更精确的特征表示[13-14]；辅助特征使用其他辅助信息(如语义属性)学习与表示特征[15]；视频特征是从多个图像帧中学习视频的特征表示，用于视频中的行人重识别[16]。

全局特征指学习每个图像的全局特征，其仅利用整个图像进行特征提取。随着深度神经网络应用于行人重识别，基于深度学习的全局特征学习已成为提取特征向量的主要策略[17]。为提取更有用的全局特征向量，身份判别嵌入模型(ID-discriminative Embedding，IDE)[1]将行人重识别视为一个多类分类问题，每个身份被视为一个不同的类。近年来，研究者们为行人重识别设计了多种用于全局特征表示的深度网络，以达到更优的行人重识别性能[2-3,18]。

本文将利用差异注意力信息增强全局特征向量的表示效果和识别能力。差异注意力不局限于全局特征表示方法，它适用于任何类型的行人重识别特征表示学习模型。

2.3　行人重识别的注意力机制

注意力方法通过关注特征向量中的重要特征抑制不相关特征，使注意力可适应复杂的任务需求。Wang 等[5]和 Yang等[6]在注意力模块中设置卷积层以获得更大的感受野。卷积块注意力模块[7]在空间特征和通道特征上利用卷积层和一个共享

的多层感知机(Multilayer Perceptron，MLP)学习空间和通道注意图。其他相关工作将人类语义的外部线索视为注意力，或将其作为辅助信息来指导注意力的学习[8-9,19-20]。

然而，上述方法仅利用了单个图像的特征生成相应的注意力信息。为进一步使用两个不同图像特征向量之间的差异信息，本文设计了差异注意力模块生成差异注意力信息，为距离函数提供更具区分度的注意力，以获得更好的行人重识别性能。

3 基于城市信息单元和差异注意力的多层行人重识别技术

本节将介绍基于城市信息单元和差异注意力的多层行人重识别技术。第 3.1 小节讨论行人重识别技术与城市信息单元的深度融合；在回顾广泛使用的有监督的行人重识别框架后，第 3.2 小节提出差异注意力模式；第 3.3 小节详细描述差异注意力模块；第 3.4 小节介绍整个差异注意力框架以及两种不同的训练策略。

3.1 基于城市信息单元的多层行人重识别

在智慧城市系统中，行人重识别任务具有重要的实践意义与应用价值。然而，在解决具体的实践问题上，单纯的行人重识别技术还存在盲点。为提高行人重识别技术的实用性，本文将行人重识别技术与多级城市信息单元深度融合，形成相似的层次结构，构建解决实际问题的概念模型，使得行人重识别技术能够满足智慧城市场景下的多层次行人重识别任务需求。

行人重识别任务可被看作安全领域的一项多层次的复杂任务，不同的行人重识别任务之间，可通过共同/不同的查询子集/图库子集形成行人重识别任务的层级关系。城市信息单元的层次结构类似。图 1 展示了城市信息单元与多级行人重识别相似的层次结构，从下到上依次为网格、区域、街道、行政区域和城市，上级城市信息单元包含下级城市信息单元，同一级别的城市信息单

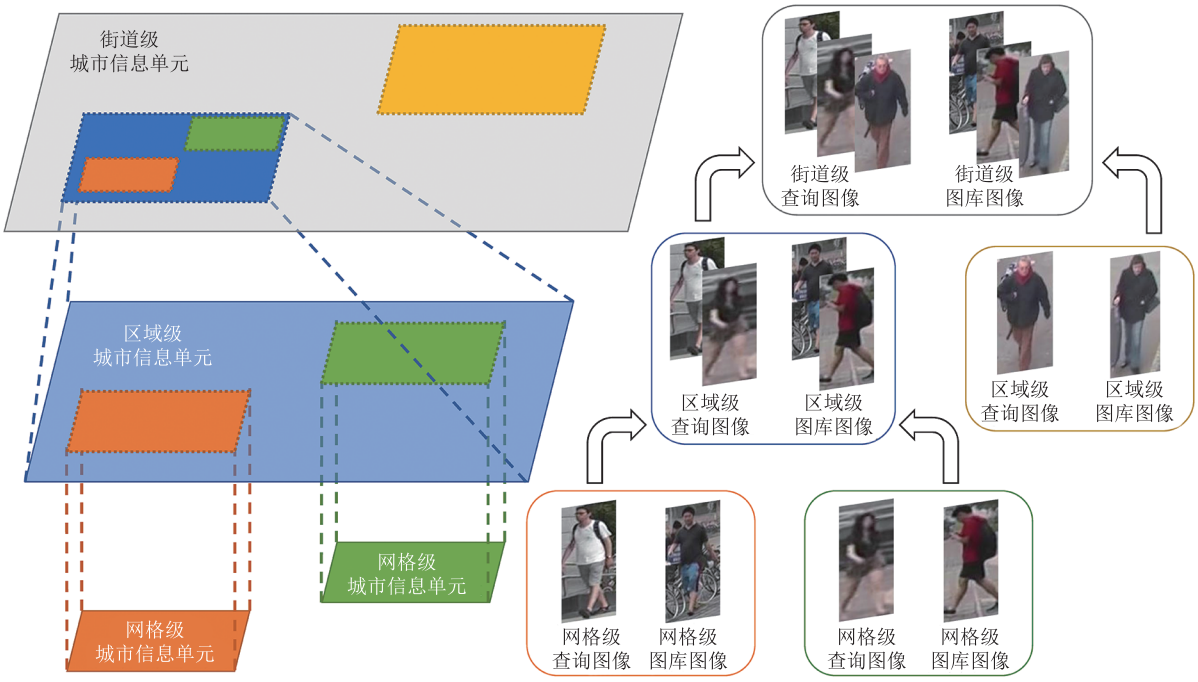


图 1 行人重识别任务与城市信息单元的层次结构

Fig. 1 The hierarchical architecture of Person Re-ID tasks and urban information units

元相互独立。同样地，同级城市信息单元需要执行的人员识别任务也是独立的，高级的城市信息单元对应的行人重识别任务包括其所有的下级城市信息单元的行人重识别任务，而最低级的城市信息单元对应的任务也是最基础的任务。

基于上述层次结构，城市信息单元可作为解决实际问题的概念模型。选择不同层级的城市信息单元，根据其包含的政府数据和社会传感器数据，即可确定具体需要执行行人重识别任务的查询图集和图库图集，从而明确地执行具体的行人重识别任务，生成查询结果以组成最终的任务输出。

将行人重识别技术与城市信息单元深度融合，可明确行人重识别任务在智慧城市等实际应用场景中的概念模型，满足多样的多级行人重识别任务需求。此外，基于城市信息单元的多级行人重识别，还可更进一步解决行人跟踪等其他与行人重识别相关的问题。

3.2　差异注意力

行人重识别旨在从预定义的图库中查找与给定的查询图像最相似的图像。一般地，通过深度学习方法进行有监督的行人重识别包括 3 个步骤：(1)提取训练数据集(通常基于 ResNet-50 骨干网络[17])的图像特征向量，并训练深度模型；(2)使用(1)中训练的模型提取查询图像和图库中所有图像的特征向量；(3)计算查询图像特征向量与图库图像特征向量之间的距离(或相似性)，并对距离矩阵进行排序，以生成行人重识别查询结果。

给定两个图像特征向量 $\boldsymbol{x}_1,\boldsymbol{x}_2\in\mathbb{R}^p$，其欧氏距离和余弦距离计算如下：

$$D_{\text{Euclidean}}(\boldsymbol{x}_1,\boldsymbol{x}_2)\downarrow=\|\boldsymbol{x}_1-\boldsymbol{x}_2\|_2=\sqrt{\sum_i^p\left(\boldsymbol{x}_1^{(i)}-\boldsymbol{x}_2^{(i)}\right)^2}\quad(1)$$

$$D_{\text{cosine}}(\boldsymbol{x}_1,\boldsymbol{x}_2)\uparrow=\frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2}\quad(2)$$

其中，p 为两个向量的维数；$\boldsymbol{x}^{(i)}$ 为特征向量中第 i 个特征；↑表示距离随图像相似度的增加而增加；↓表示距离随图像相似度的增加而减少。

对于余弦距离，给定任何图像的特征向量 $\boldsymbol{x}$，通过标准化使得 $\|\boldsymbol{x}\|_2=\sum_i^p\left(\boldsymbol{x}^{(i)}\right)^2=1$。余弦距离也可看作两个特征向量之差的函数。

$$\begin{aligned}D_{\text{cosine}}(\boldsymbol{x}_1,\boldsymbol{x}_2)\uparrow&=\frac{\boldsymbol{x}_1\cdot\boldsymbol{x}_2}{\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2}=\boldsymbol{x}_1\cdot\boldsymbol{x}_2\\&=1-\frac{1}{2}\sum_i^p\left(\left(\boldsymbol{x}_1^{(i)}\right)^2-2\boldsymbol{x}_1^{(i)}\boldsymbol{x}_2^{(i)}+\left(\boldsymbol{x}_2^{(i)}\right)^2\right)\\&=1-\frac{1}{2}\sum_i^p\left(\boldsymbol{x}_1^{(i)}-\boldsymbol{x}_2^{(i)}\right)^2\\&=1-\frac{1}{2}D_{\text{Euclidean}}^2(\boldsymbol{x}_1,\boldsymbol{x}_2)\downarrow\end{aligned}\quad(3)$$

在许多情况下，提取鲁棒的图像特征是行人重识别任务中最重要的部分。由实验结果可知，不同类型的图像特征在不同的任务和数据集上可能具有最佳性能。当深度网络模型的参数固定时，图像特征将失去针对不同任务的灵活性和鲁棒性。因此，本文在解决行人重识别任务时，需要提供能够提取各种特征的深度网络，当计算不同行人重识别任务中图像特征之间的距离时，选择合适的特征就变得尤为重要。使用上述差异注意力模式，根据特征向量的差异注意力对特征进行加权，距离函数只需计算两个特征向量之间的有用特征差异，就可计算出更具辨别力的距离矩阵。

3.3　差异注意力模块

差异注意力模块是差异注意力框架中的核心组件，其结构如图 2 所示，差异注意力模块包括输入变换、聚合卷积、多层感知机和输出变换等组件。

差异注意力模块的输入是一对图像特征向量，其中，查询图像的特征向量为 $\boldsymbol{x}_q\in\mathbb{R}^p$，图库图像的特征向量 $\boldsymbol{x}_g\in\mathbb{R}^p$。首先，输入这两个特征向量到输入变换模块生成差异特征向量 $\boldsymbol{x}_d\in\mathbb{R}^p$；然后，这 3 个向量通过聚合卷积层获取特征信息向量 $\boldsymbol{x}_f\in\mathbb{R}^p$。在聚合卷积步骤中，特征信息向量 $\boldsymbol{x}_f$ 的计算公式为：

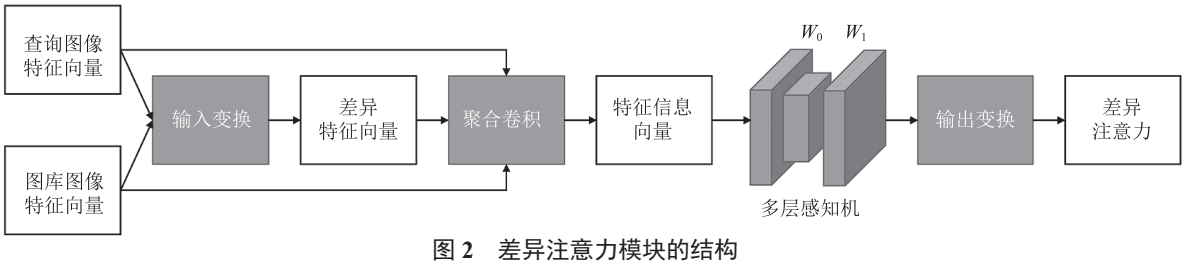


图 2 差异注意力模块的结构

Fig. 2 The structure of our diff attention module

$$\begin{aligned} \boldsymbol{x}_f &= f^{1\times1}\left(\left[\mu\left(\boldsymbol{x}_q,\boldsymbol{x}_g\right);\boldsymbol{x}_q;\boldsymbol{x}_g\right]\right) \\ &= f^{1\times1}\left(\left[\boldsymbol{x}_d;\boldsymbol{x}_q;\boldsymbol{x}_g\right]\right) \end{aligned} \tag{4}$$

其中，$f^{1\times1}$ 为 1×1 的聚合卷积层；μ 为输入变换模块，输入变换过程为：$\boldsymbol{x}_g$ 减去 $\boldsymbol{x}_q$，再对运算结果求绝对值，以获得对称的差异注意力效果。

基于特征信息向量 $\boldsymbol{x}_f$，利用 MLP 计算差异注意力图 $M_d \in \mathbb{R}^p$。差异注意力图的计算公式如下：

$$M_d = \sigma\left(MLP\left(\boldsymbol{x}_f\right)\right) = \sigma\left(W_1\left(ReLU\left(W_0\left(\boldsymbol{x}_f\right)\right)\right)\right) \tag{5}$$

其中，σ 为输出变换的 sigmoid 函数；W_0 和 W_1 为多层感知机中的参数，$W_0 \in \mathbb{R}^{p/ratio\times p}$，$W_1 \in \mathbb{R}^{p\times p/ratio}$；*ReLU* 为多层感知机中使用的 ReLU 激活函数。通过差异注意力模块计算出一维的差异注意力图 M_d，距离函数的差异注意过程可以表述为：

$$\begin{aligned} D_{DA}\left(\boldsymbol{x}_q,\boldsymbol{x}_g\right) &= \sqrt{\sum_i^p\left[M_d^{(i)}\left(\boldsymbol{x}_q^{(i)}-\boldsymbol{x}_g^{(i)}\right)\right]^2} \\ &= \sqrt{\sum_i^p\left(M_d^{(i)}\boldsymbol{x}_q^{(i)}-M_d^{(i)}\boldsymbol{x}_g^{(i)}\right)^2} \\ &= D_{\text{Euclidean}}\left(M_d\otimes\boldsymbol{x}_q, M_d\otimes\boldsymbol{x}_g\right) \\ &= D_{\text{Euclidean}}\left(\boldsymbol{x}_q', \boldsymbol{x}_g'\right) \end{aligned} \tag{6}$$

其中，⊗为元素乘法。上述过程表明，使用差异注意力机制的 $\boldsymbol{x}_q$ 和 $\boldsymbol{x}_g$ 之间的距离可被视为 $\boldsymbol{x}_q'$ 和 $\boldsymbol{x}_g'$ 之间的距离，$\boldsymbol{x}_q' = M_d\otimes\boldsymbol{x}_q$，$\boldsymbol{x}_g' = M_d\otimes\boldsymbol{x}_g$。

3.4 差异注意力框架

为最终实现差异注意力，本文设计了用于行人重识别的差异注意力框架，结构如图 3 所示，其主要结构包括骨干网络（BagTricks 或 AGW）、差异注意力模块及距离函数。

首先，利用骨干网络提取图像的深度特征向量。然后，差异注意力框架中的特征向量将被成对地发送到差异注意力模块，以生成每对图像之间的差异注意力图，再将差异注意力与原始的特征向量相乘。在训练阶段，训练批次中每个图像的特征向量与同一批次中的所有其他向量互相配对，以计算差异注意力图；在推理阶段，查询图像的特征向量和图库图像的特征向量自然配对。最后，可以通过距离函数计算图像对之间的距离，以计算损失，从而训练深度模型或得到行人重识别结果。

为了使差异注意力框架适用于多种经过训练的深度模型，本文还提出了联合训练和单独训练两种训练策略。联合训练通常用于训练新的深度

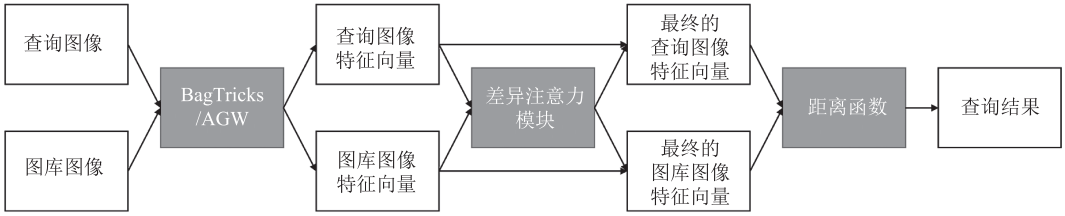


图 3 差异注意力框架的结构

Fig. 3 The structure of our diff attention framework

网络，单独训练则更适用于微调已经训练过的深度模型。

联合训练指一起训练所有的模型，即同时训练骨干模型和差异注意力模块。该训练策略有助于训练适应差异注意力的骨干网络。在联合训练开始前，通常利用 ImageNet 数据集预训练骨干模型，并且随机初始化差异注意力模块。联合训练中涉及的训练超参数与仅训练骨干模型的参数相同，并采用骨干模型 BagTricks[4]和 AGW[21]论文中所使用的损失函数，损失函数及其参数保持不变。联合训练使用 ID 损失 L_{ID} 和标签平滑技术[22]、三元组损失 $L_{\mathrm{BHTriplet}}$[23]和中心损失 L_{Center}[24]来训练所有的模型。对于 AGW 骨干模型，将使用其加权正则化三元组损失[21]。

联合训练的损失函数公式如下：

$$L_{\mathrm{Joint}}=\alpha L_{\mathrm{ID}}+\beta L_{\mathrm{BHTriplet}}+\gamma L_{\mathrm{Center}} \tag{7}$$

其中，α 为损失函数 L_{ID} 的权重；β 为损失函数 $L_{\mathrm{BHTriplet}}$ 的权重，γ 为损失函数 L_{Center} 的权重。在联合训练中，ε=0.1、m=0.3、$\alpha=\beta$=1、γ=0.000 5。

单独训练指微调现有的训练过的骨干模型。训练模型的超参数可能与仅训练骨干模型时使用的参数不同。该训练策略可以大大缩短训练时间和训练成本，有助于快速找到差异注意力框架的最佳超参数。由于不再训练骨干模型，联合训练使用的损失函数中只有三元组损失具有意义，ID 损失与中心损失不再发生改变。因此，单独训练可仅使用三元组损失训练差异注意力模块。

单独训练的损失函数公式如下：

$$L_{\mathrm{Separate}}=\beta L_{\mathrm{BHTriplet}} \tag{8}$$

其中，β 是三元组损失的权重。由于骨干模型进行了预训练，因此，三元组损失较小。为保证模型训练收敛与训练效率，进行相关实验测试。测试结果表明，当单独训练中的 β=10 时，可以增强损失函数的训练效果。

为增强三元组损失的效果，在计算三元组损失时，使用 softplus 函数而非 hinge 函数，这被称为 soft-margin 方法[23]。

4　实　验

本节将评估差异注意力框架的行人重识别性能。第 4.1 小节将介绍实验中使用的数据集；第 4.2 小节将列出所有的实现细节；第 4.3 小节将验证差异注意力模块的效果；第 4.4 小节将对差异注意力框架所涉及的超参数进行讨论；第 4.5 小节将差异注意力框架与其他最先进的有监督的行人重识别方法进行对比；第 4.6 小节主要介绍基于城市信息单元的安防监控识别系统的具体应用。

4.1　数据集

本实验使用了 3 个著名的基于图像的行人重识别数据集：Market-1501[25]、CUHK03[26]和 MSMT17[27]。其中，Market-1501 包括 32 668 个有标签的行人边界框，每个边界框由 DPM 模型[28]检测而来，每个身份至少由 2 个摄像头捕捉，数据集包含 6 个摄像机捕捉到的 1 501 个身份；CUHK03 包含 1 360 名行人的 13 164 张图片，数据集由 6 个摄像头捕获，每个身份由 2 个不相交的摄像头进行观察；MSMT17 是一个新的多场景多时间的行人重识别数据集，尽可能地模拟了真实场景，其数据由部署在校园内的 15 个摄像头网络进行收集，该数据集包括 4 101 名行人的 126 441 个边界框。

4.2　训练设置

差异注意力框架的骨干模型是 AGW 基线网络[21]和 BagTricks 强基线[4]，它们均使用经 ImageNet 预训练后的 ResNet-50[17]作为骨干网络。

本实验中所有的模型训练硬件为 NVIDIA GeForce RTX 3080 Ti。所有图像的尺寸被调整为 256×128，每张图像填充 10 个像素并被随机裁剪。此外，本模型还使用了一些被广泛使用的图像增强方法：随机水平翻转和随机擦除增强[29]，

翻转概率 $p=0.5$。

为计算 ID 损失，本实验在骨干模型后添加了一个无偏差的全连接层。该层的输出维度设置为训练集中的身份数。由于 GPU 显存容量的限制，批次大小被限制为 64，并设置 $P=16$，$K=4$。优化中心损失的中心参数的算法是 SGD。

训练使用的优化模型算法是 Adam，权重衰减为 5×10^{-4}。联合训练共设置 120 个训练回合，初始学习率为 3.5×10^{-4}，在前 10 个回合预热学习率[30]，在第 40 个和第 70 个回合学习率降低为原来的 1/10。对于单独训练，只训练 60 个回合，初始学习率设置为 0.05，每 20 个回合降低一次学习率。

对于差异注意力模块，输入变换是带绝对值的减法。当骨干模型为 AGW 时，MLP 比率设置为 4，当骨干模型为 BagTricks 时，MLP 比率设置为 512。

本文使用累积匹配特性、平均准确率和平均逆负惩罚 3 个评估指标评估差异注意力框架的性能。值得注意的是，本实验未使用重排序技术[31]。

4.3 差异注意力的效果

本节将展示两种不同训练策略下的差异注意力框架的实验结果。本实验使用单独训练的策略，以寻求差异注意力模块的最佳参数。

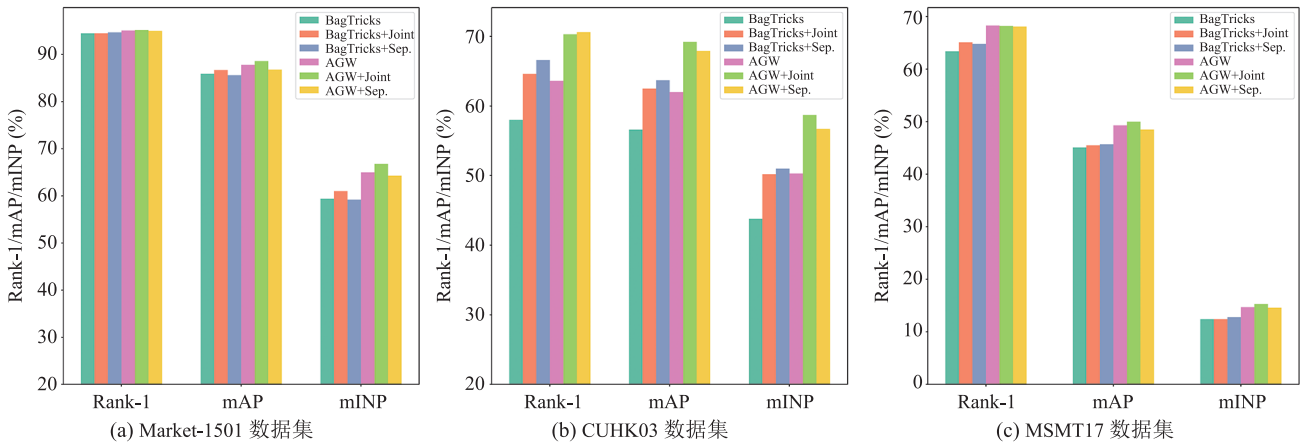
如表 1 和图 4 所示，在 CUHK03 数据集上，差异注意力框架与联合训练分别获得了 64.6% 和 70.3% 的 Rank-1 准确度、62.5% 和

表 1 差异注意力框架的性能

Table 1 The performance of our diff attention framework

策略	Market-1501			CUHK03			MSMT17		
	Rank-1 (%)	mAP (%)	mINP (%)	Rank-1 (%)	mAP (%)	mINP (%)	Rank-1 (%)	mAP (%)	mINP (%)
BagTricks[4]	94.5	85.9	59.4	58.0	56.6	43.8	63.4	45.1	12.4
BagTricks[4]+Joint	94.5	86.7	61.0	64.6	62.5	50.2	65.1	45.5	12.4
BagTricks[4]+Sep.	94.7	85.6	59.2	66.6	63.7	51.0	64.8	45.7	12.8
AGW[21]	95.1	87.8	65.0	63.6	62.0	50.3	68.3	49.3	14.7
AGW[21]+Joint	95.2	88.6	66.8	70.3	69.2	58.7	68.2	50.0	15.3
AGW[21]+Sep.	95.0	86.8	64.3	70.6	67.9	56.7	68.1	48.5	14.6

注：“Joint”代表使用联合训练策略；“Sep.”代表使用单独训练策略



(a) Market-1501 数据集　(b) CUHK03 数据集　(c) MSMT17 数据集

图 4 差异注意力框架的性能

Fig. 4 The performance of our diff attention framework

69.2% 的 mAP、50.2% 和 58.7% 的 mINP；单独训练也获得了良好的结果：66.6% 和 70.6% 的 Rank-1 准确度、63.7% 和 67.9% 的 mAP、51.0% 和 56.7% 的 mINP。在 Market-1501 数据集上，本实验使用 AGW 主干模型的框架取得了 95.2% 的 Rank-1 准确度、88.6% 的 mAP 和 66.8% 的 mINP，高于原始 AGW 基线模型的性能。在 MSMT17 数据集上，使用 AGW 模型的训练结果为 68.2% 的 Rank-1 准确度、50.0% 的 mAP 和 15.3% 的 mINP。

4.4　消融实验

本节将通过实验对差异注意力模块中的超参数进行讨论——在所有的消融实验中，将 BagTricks 和 AGW 作为骨干网络，使用单独训练的策略，分别对不同的超参数进行实验，确定模型超参数的最优值。

4.4.1　输入变换

本文比较了 3 种输入变换方法(减法、减法后平方和减法后绝对值)的影响。在这些输入变换的消融实验中，当 AGW 作为主干模型时，MLP 比率固定为 4；当 BagTricks 作为主干模型时，MLP 比率固定为 512。

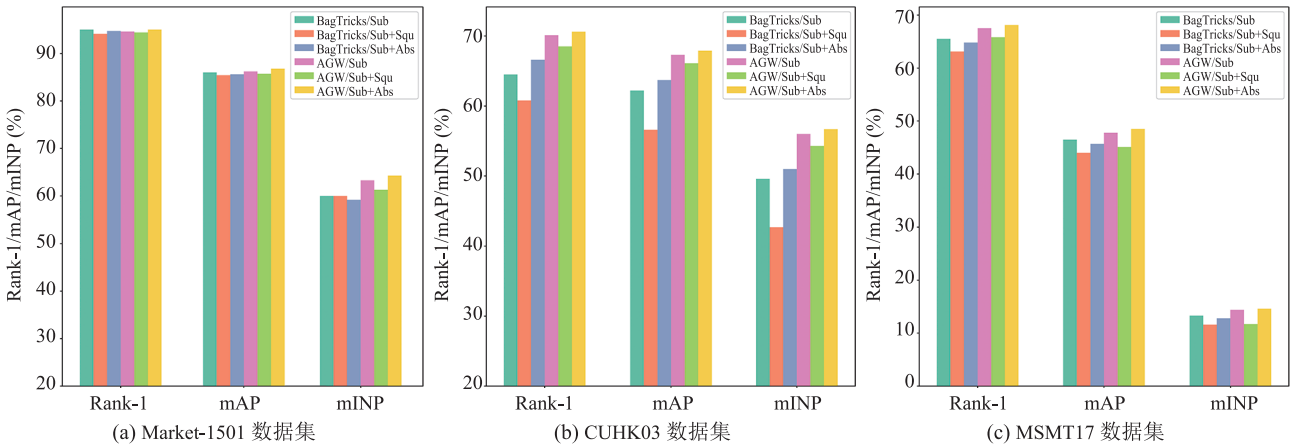
表 2 和图 5 的输入变换实验结果显示了不同输入变换对模型性能的影响。由此可知，依次进行减法运算和取绝对值运算的输入变换取得了最好的性能，其在 AGW 模型或 CUHK03 数据集上均实现了最佳性能。与其他两种输入变换相比，仅进行减法运算的性能较差。

表 2　不同输入变换的影响

Table 2　The impact of different input transforms

骨干模型	输入变换	Market-1501			CUHK03			MSMT17		
		Rank-1 (%)	mAP (%)	mINP (%)	Rank-1 (%)	mAP (%)	mINP (%)	Rank-1 (%)	mAP (%)	mINP (%)
BagTricks[4]	Sub	95.0	86.0	60.0	64.5	62.2	49.6	65.5	46.5	13.3
BagTricks[4]	Sub＋Squ	94.1	85.4	60.0	60.8	56.6	42.7	63.1	44.0	11.6
BagTricks[4]	Sub+Abs	94.7	85.6	59.2	66.6	63.7	51.0	64.8	45.7	12.8
AGW[21]	Sub	94.6	86.2	63.3	70.1	67.3	56.0	67.5	47.8	14.4
AGW[21]	Sub＋Squ	94.4	85.7	61.3	68.5	66.1	54.3	65.8	45.1	11.7
AGW[21]	Sub＋Abs	95.0	86.8	64.3	70.6	67.9	56.7	68.1	48.5	14.6

注：“Sub”代表仅进行减法运算；“Sub＋Squ”代表依次进行减法运算和平方运算；“Sub＋Abs”代表依次进行减法运算和取绝对值运算



(a) Market-1501 数据集　(b) CUHK03 数据集　(c) MSMT17 数据集

图 5　不同输入变换的影响

Fig. 5　The impact of different input transforms

4.4.2 MLP 比率

MLP 比率是差异注意力模块的核心超参数，本节通过实验比较了不同的 MLP 比率对模型性能的影响。基于第 4.4.1 小节的实验结果，在测试时将输入变换固定为带绝对值的减法。图 6 为不同 MLP 比率的影响，当使用 AGW 作为主干模型时，将 MLP 比率设置为 4，通常可实现最佳性能；若使用 BagTricks，那么就将比率设置为 512。

4.5 与先进方法的对比

本文将其他先进方法分为全局特征和其他两种不同的类型，并与差异注意力框架进行比较，结果如表 3～5 所示。由表 3～5 可知，差异注意力方法的 mAP 和 Rank-1 准确度均较为优异。

4.6 基于城市信息单元的安防监控识别系统

本文将行人重识别技术与城市信息单元深度融合，基于自建数据集，实现了基于城市信息单元的安防监控识别系统，如图 7 所示。用户上传待查询的行人图像到该系统后，系统对行人图像

表 3 在 Market-1501 上与其他最先进方法的比较结果

Table 3 Comparison results with other state-of-the-art methods on Market-1501

类型	方法	年份	Market-1501	
			Rank-1 (%)	mAP (%)
全局特征	IDE[1]	2017	79.5	59.9
全局特征	SVDNet[18]	2017	82.3	62.1
全局特征	TriNet[23]	2017	84.9	69.1
全局特征	BagTricks[4]	2019	94.5	85.9
全局特征	AGW[21]	2021	95.1	87.8
其他	PCB[2]	2018	93.8	81.6
其他	Mancs[32]	2018	93.1	82.3
其他	IANet[33]	2019	94.4	83.1
其他	DGNet[34]	2019	94.8	86.0
其他	PyrNet[35]	2019	93.6	81.7
其他	Auto-ReID[36]	2019	94.5	85.1
其他	OSNet[37]	2019	94.8	86.7
其他	Circle[38]	2020	94.2	84.9
其他	HOReID[39]	2020	94.2	84.9
其他	TransReID[40]	2021	95.0	88.2
AGW＋差异注意力＋联合训练			95.2	88.6
AGW＋差异注意力＋单独训练			95.0	86.8

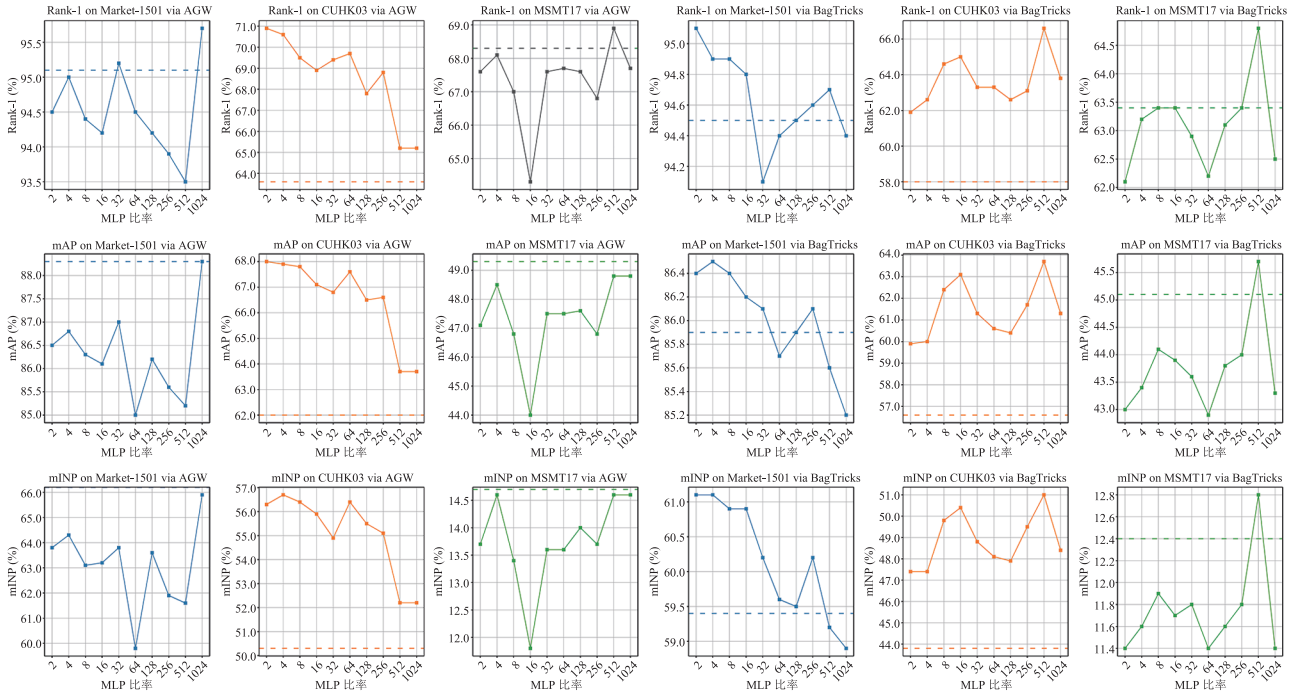


图 6 不同 MLP 比率的影响

Fig. 6 The impact of different MLP ratios

表 4　在 CUHK03 上与其他最先进方法的比较结果

Table 4　Comparison results with other state-of-the-art methods on CUHK03

类型	方法	年份	CUHK03	
			Rank-1 (%)	mAP (%)
全局特征	SVDNet[18]	2017	41.5	37.3
全局特征	BagTricks[4]	2019	58.0	56.6
全局特征	AGW[21]	2021	63.6	62.0
其他	PCB[2]	2018	63.7	57.5
其他	Mancs[32]	2018	65.5	60.5
其他	MGN[41]	2018	66.8	66.0
其他	DGNet[34]	2019	65.6	61.1
其他	PyrNet[35]	2019	68.0	63.8
AGW＋差异注意力＋联合训练			70.3	69.2
AGW＋差异注意力＋单独训练			70.6	67.9

表 5　在 MSMT17 上与其他最先进方法的比较结果

Table 5　Comparison results with other state-of-the-art methods on MSMT17

类型	方法	年份	MSMT17	
			Rank-1 (%)	mAP (%)
全局特征	BagTricks [4]	2019	63.4	45.1
全局特征	AGW[21]	2021	68.3	49.3
其他	PCB[2]	2018	68.2	40.4
其他	IANet[33]	2019	75.5	46.8
其他	CBN[42]	2020	72.8	42.9
AGW＋差异注意力＋联合训练			68.2	50.0
AGW＋差异注意力＋单独训练			68.1	48.5

进行图像增强，并利用行人重识别深度模型进行特征提取。识别系统将依次对提取的行人图像特征与选定的城市信息单元中对应的图库图像特征进行相似度计算，并根据相似度排序生成识别结果序列。识别系统还能综合行人重识别结果与城市信息单元中的位置数据，利用地图组件生成待查询行人的轨迹。实验结果表明，本文基于城市信息单元的安防监控识别系统识别精度高，生成识别结果速度较快，轨迹展示效果直观明显。

5　结　论

为提高行人重识别技术在智慧城市等现实场景中的应用能力，本文提出将行人重识别技术与城市信息单元进行多层次深度融合。在行人重识别的过程中，特征差异具有重要作用。因此，本文提出了差异注意力的概念，主张利用差异注意力模块实现深度特征的差异注意力机制；并提出了差异注意力框架，使得差异注意力模块适用于多种深度特征模型。此外，本文还提出两种不同的训练策略(联合训练和单独训练)，以训练差异注意力框架，快速找到能够获得最佳性能的参

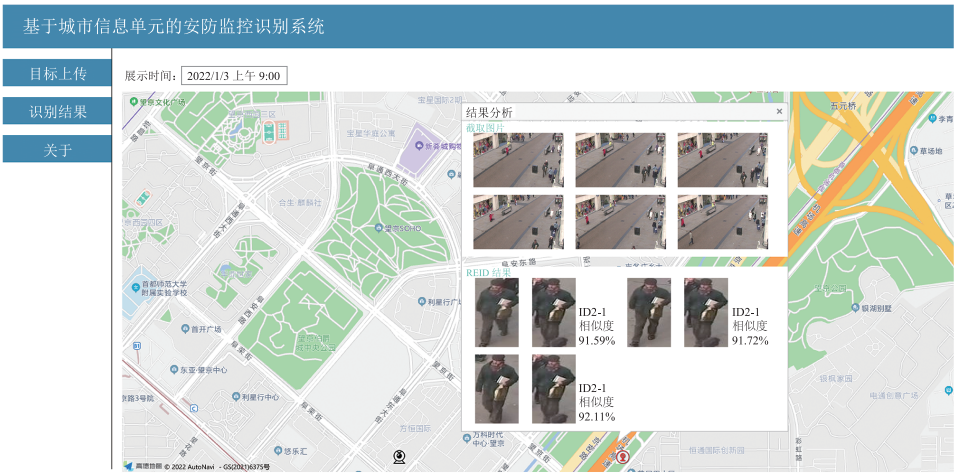


图 7　基于城市信息单元的安防监控识别系统

Fig. 7　The identification system based on urban information unit

数。在 Market-1501、CUHK03 和 MSMT17 上，与其他先进的行人重识别方法相比，差异注意力框架行人重识别性能较为优异。最后，期望本研究能为行人重识别技术在现实场景中的广泛应用做出贡献。

参 考 文 献

[1] Zheng L, Zhang HH, Sun SY, et al. Person re-identification in the wild [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1367-1376.

[2] Sun YF, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C] // Proceedings of the European Conference on Computer Vision, 2018: 480-496.

[3] Qian XL, Fu YW, Jiang YG, et al. Multi-scale deep learning architectures for person re-identification [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 5399-5408.

[4] Luo H, Gu YZ, Liao XY, et al. Bag of tricks and a strong baseline for deep person re-identification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019: 1487-1495.

[5] Wang F, Jiang MQ, Qian C, et al. Residual attention network for image classification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3156-3164.

[6] Yang F, Yan K, Lu SJ, et al. Attention driven person re-identification [J]. Pattern Recognition, 2019, 86: 143-155.

[7] Woo S, Park J, Lee J, et al. CBAM: convolutional block attention module [C] // Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[8] Xu J, Zhao R, Zhu F, et al. Attention-aware compositional network for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 2119-2128.

[9] Zhao HY, Tian MQ, Sun SY, et al. Spindle Net: person re-identification with human body region guided feature decomposition and fusion [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1077-1085.

[10] 徐志杰. 面向人群流量预测的城市信息单元画像建模 [D]. 北京: 北京交通大学, 2021.
Xu ZJ. Urban information unit profiling for crowd flow prediction [D]. Beijing: Beijing Jiaotong University, 2021.

[11] 李擎, 胡伟阳, 李江昀, 等. 基于深度学习的行人重识别方法综述 [J]. 工程科学学报, 2022, 44(5): 920-932.
Li Q, Hu WY, Li JY, et al. A survey of person re-identification based on deep learning [J]. Chinese Journal of Engineering, 2022, 44(5): 920-932.

[12] 杨永胜, 邓淼磊, 李磊, 等. 基于深度学习的行人重识别综述 [J]. 计算机工程与应用, 2022, 58(9): 51-66.
Yang YS, Deng ML, Li L, et al. Overview of pedestrian re-identification based on deep learning [J]. Computer Engineering and Applications, 2022, 58(9): 51-56.

[13] Zhao LM, Li X, Zhuang YT, et al. Deeply-learned part-aligned representations for person re-identification [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 3219-3228.

[14] Yao HT, Zhang SL, Hong RC, et al. Deep representation learning with part loss for person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28(6): 2860-2871.

[15] Su C, Zhang SL, Xing JL, et al. Deep attributes driven multi-camera person re-identification [C] // Proceedings of the European Conference on Computer Vision, 2016: 475-491.

[16] Dai J, Zhang PP, Wang D, et al. Video person re-

identification by temporal residual learning [J]. IEEE Transactions on Image Processing, 2018, 28(3): 1366-1377.

[17] He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

[18] Sun YF, Zheng L, Deng WJ, et al. SVDNet for pedestrian retrieval [C] // Proceedings of the IEEE International Conference on Computer Vision, 2017: 3800-3808.

[19] Song CF, Huang Y, Ouyang WL, et al. Mask-guided contrastive attention model for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 1179-1188.

[20] Suh Y, Wang JD, Tang SY, et al. Part-aligned bilinear representations for person re-identification [C] // Proceedings of the European Conference on Computer Vision, 2018: 402-419.

[21] Ye M, Shen JB, Lin GJ, et al. Deep learning for person re-identification: a survey and outlook [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(6): 2872-2893.

[22] Christian S, Vincent V, Sergey I, et al. Rethinking the inception architecture for computer vision [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.

[23] Alexander H, Lucas B, Bastian L. In defense of the triplet loss for person re-identification [Z/OL]. arXiv Preprint, arXiv: 1703.07737, 2017.

[24] Wen YD, Zhang KP, Li ZF, et al. A discriminative feature learning approach for deep face recognition [C] // Proceedings of the European Conference on Computer Vision, 2016: 499-515.

[25] Zheng L, Shen LY, Tian L, et al. Scalable person re-identification: a benchmark [C] // Proceedings of the IEEE International Conference on Computer Vision, 2015: 1116-1124.

[26] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 152-159.

[27] Wei LH, Zhang SL, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 79-88.

[28] Felzenszwalb PF, Girshick RB, McAllester D, et al. Object detection with discriminatively trained part-based models [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 32(9): 1627-1645.

[29] Zhong Z, Zheng L, Kang GL, et al. Random erasing data augmentation [C] // Proceedings of the AAAI Conference on Artificial Intelligence, 2020: 13001-13008.

[30] Fan X, Jiang W, Luo H, et al. SphereReID: deep hypersphere manifold embedding for person re-identification [J]. Journal of Visual Communication of Image Representation, 2019, 60: 51-58.

[31] Zhong Z, Zheng L, Cao DL, et al. Re-ranking person re-identification with k-reciprocal encoding [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1318-1327.

[32] Wang C, Zhang Q, Huang C, et al. Mancs: a multi-task attentional network with curriculum sampling for person re-identification [C] // Proceedings of the European Conference on Computer Vision, 2018: 365-381.

[33] Hou RB, Ma BP, Chang H, et al. Interaction-and-aggregation network for person re-identification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9317-9326.

[34] Zheng ZD, Yang XD, Yu ZD, et al. Joint discriminative and generative learning for person re-identification [C] // Proceedings of the IEEE/

CVF Conference on Computer Vision and Pattern Recognition, 2019: 2138-2147.

[35] Martinel N, Luca Foresti G, Micheloni C. Aggregating deep pyramidal representations for person re-identification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019: 1544-1554.

[36] Quan RJ, Dong XY, Wu Y, et al. Auto-reID: searching for a part-aware ConvNet for person re-identification [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3750-3759.

[37] Zhou KY, Yang YX, Cavallaro A, et al. Omni-scale feature learning for person re-identification [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 3702-3712.

[38] Sun YF, Cheng CM, Zhang YH, et al. Circle loss: a unified perspective of pair similarity optimization [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6398-6407.

[39] Wang GA, Yang S, Liu HY, et al. High-order information matters: learning relation and topology for occluded person re-identification [C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6449-6458.

[40] He ST, Luo H, Wang PC, et al. TransReID: transformer-based object re-identification [C] // Proceedings of the IEEE/CVF International Conference on Computer Vision, 2020: 15013-15022.

[41] Wang GS, Yuan YF, Chen X, et al. Learning discriminative features with multiple granularities for person re-identification [C] // Proceedings of the 26th ACM International Conference on Multimedia, 2018: 274-282.

[42] Zhuang ZJ, Wei LH, Xie LX, et al. Rethinking the distribution gap of person re-identification with camera-based batch normalization [C] // Proceedings of the European Conference on Computer Vision, 2020: 140-157.