

基于多视角跨模态的电力现场作业 行人重识别网络架构技术研究

张森¹, 张颀², 王尧³, 刘锦隆², 闫斌⁴, 尚赵伟¹

(1. 重庆大学计算机学院, 重庆 400000; 2. 国网四川省电力公司, 四川 成都 610041;

3. 国网四川省电力公司凉山供电公司, 四川 西昌 615000; 4. 电子科技大学自动化学院, 四川 成都 611731)

摘要:可见光到红外光跨模态行人重识别目的是实现在白天和夜间环境下对行人身份的识别判断,在视频监控领域具有重要研究价值。因可见光和红外光成像原理的不同,给跨模态重识别问题带来了挑战。设计了一种新的网络结构,用于缓解模态间数据差异,提高行人重识别模型的精度。网络结构分为两部分:基于注意力的模态迁移模块嵌入特征网络的输入级,可缩小跨模态差异;基于分块的多粒度特征分解模块,同时考虑整体信息和局部信息并提高了有效信息的利用率。在公开数据集 SYSU-MM01 上,所提方法的累计匹配特性指标的 rank1 达到了 56.45%,平均精确度指标达到了 53.52%,比当前最佳方法(XIV, AAAI-2020)分别提高了 6.53% 和 2.79%,有效提高了可见光到红外光跨模态行人重识别的性能。

关键词:行人重识别; 跨模态; 注意力; 多粒度特征

中图分类号: TP394.1 **文献标志码:** A **文章编号:** 1003-6954(2020)06-0006-05

Research on Network Structure of Pedestrian Recognition for Power Operation Field Based on Multi-view Cross-modal Image Processing

Zhang Sen¹, Zhang Jie², Wang Yao³, Liu Jinglong², Yan Bin⁴, Shang Zhaowei¹

(1. College of Computer Science, Chongqing University, Chongqing 400000, China;

2. State Grid Sichuan Electric Power Company, Chengdu 610041, Sichuan, China;

3. State Grid Liangshan Electric Power Supply Company, Xichang 615000, Sichuan, China; 4. School of Automation Engineering, University of Electronic Science and Technology, Chengdu 611731, Sichuan, China)

Abstract: The purpose of cross-modal pedestrian recognition from visible to infrared light is to realize the identification and judgment of pedestrian identity in day and night environments, which is of great research value in the field of video surveillance. Due to the different imaging principles of visible light and infrared light, cross-modal recognition is a challenge. A new network structure is designed to alleviate the data difference between modals and improve the accuracy of pedestrian recognition model. The proposed network structure is divided into two parts: attention-based modal transfer module embedded in the input stage of the feature network, which can reduce the difference across modals, and block-based multi-granularity feature decomposition module, which can consider both global information and local information and improve the utilization rate of effective information. The experimental results on the open data set SYSU-MM01 show that, the rank 1 of the cumulative matching characteristic (CMC) index and the mean average precision (mAP) index of the proposed method reaches 56.45% and 53.52%, respectively, which are 6.53% and 2.79% higher than the current best method (XIV, AAAI-2020), and effectively improves the performance of cross-modal pedestrian recognition from visible to infrared light.

Key words: pedestrian recognition; cross modal; attention; multi-granularity feature

0 引言

电力是经济发展的能源基础,电力信息化对于

中国的能源安全而言无疑是一个有效的保障。近几年,视频作为安全管理和风险控制的重要手段,在电力各个业务板块得到广泛应用。行人重识别(person re-identification, ReID)是计算机视觉领域

的一个重要研究方向,要求根据某个摄像头捕捉到的行人图像在其他摄像头捕捉的图像中进行检索,以找出相同身份的行人^[1],能够根据行人的穿着、体态、发型等信息认知行人,实现对无法获取清晰拍摄人脸的行人进行跨摄像头连续跟踪,增强数据的时空连续性。随着新基建的建设快速推进,在电力行业的智慧工地、变电站监控、作业施工管控等场景中,作为人脸识别技术在身份确认方面的重要补充技术手段,行人重识别有着重要的应用价值。

在视频监控网络中,行人因受到视角、姿态变化、背景、光照、摄像头设置等多种因素影响,会导致同一个人同一摄像机的不同时刻,或不同摄像机在同时刻捕捉的图像存在很大差异,造成行人识别精度低,特别是在白天和夜晚环境下,摄像机在可见光和红外光不同模式下成像的差异更大,给跨模态(cross modal)行人重识别问题(即在可见光图像库中如何准确检索出与给定红外图像相同身份的行人)带来相当大的困难。现有研究主要集中于使用共享参数的多层卷积网络直接提取两种模态数据中的共享信息。但数据本身跨模态差异大,这些方法均仍未达到实用效果。为此提出一种新的网络结构,实验表明,可有效提取出两种模态中的共享信息,提高行人重识别精度。

1 跨模态行人重识别相关的研究

针对可见光与红外光的跨模态图像行人重识别问题,文献[2]在国际计算机视觉大会上首次发布大规模可见光-红外跨模态行人重识别数据集 SYSU-MM01,并提出深度零值填充(deep zero padding)的网络模型,开启了可见光-红外行人重识别问题的先河。文献[3]在国际人工智能联合会议上引入生成对抗网络的思想训练一个跨模态生成对抗网络模型(cGAN),使用可见光和红外光两种模态图像进行互相生成以降低模态间差异。文献[4]在国际人工智能联合会议上提出一个双流网络结构分别提取两个模态的特定特征,随后通过权重共享提取两个模态的共享特征,并于2020年^[5]使用新的基准模型对其进行改进实现了较大的精度提升。文献[6]针对跨模态问题设计了五元组损失函数,在提高同一模态类间差异的同时降低跨模态类内差异。文献[7]在国际计算机视觉与模式识别会议上使用生成

对抗的思想将两种模态图像进行互相生成,但训练不稳定且效果一般。文献[8]在美国人工智能协会2020大会上通过引入一个辅助的x模态将双模态重识别转化为可见光、红外光、x模态的三模态重识别问题,实现了较好的效果。以上研究尽管取得了一定的成就,但存在以下问题:

1)直接使用共享参数提取与模态无关的身份信息,忽视了不同模态数据本身的差异。可见光图像为红绿蓝三通道数据,包含了丰富的色彩信息,而红外图像为单通道,其灰度值反映了目标的红外辐射量,成像机理完全不同导致跨模态行人重识别领域相关研究效果较差。

2)仅考虑了图片整体特征或仅考虑局部特征,对图像信息的利用不充分。

为解决以上两个不足,下面提出的网络结构可有效提取出两种模态中的共享信息,提高跨模态行人重识别模型精度:

1)基于注意力机制设计了轻量化的模态迁移模块,可有效地在特征嵌入网络的输入级缩小跨模态差异,并且额外参数量可忽略不计。

2)设计了基于分块的多粒度特征分解模块。同时考虑输入图像的整体信息和丰富的局部信息,提高了有效信息的利用率。

3)整个模型可以端到端进行训练,并且两种模块可以独立使用,也可以与其他相关研究进行组合。

2 网络模块设计

2.1 整体网络架构

目前基于深度学习的跨模态行人重识别,多采用使用共享参数的多层卷积神经网络进行特征提取^[4-8],将可见光图像特征和红外光图像特征嵌入同一个特征空间用于后续分类(或相似度度量)。这里也采用此类网络架构,其总体框架如图1所示。对于可见光图像,首先输入到模态迁移模块,利用通道间的注意力机制将RGB 3个通道进行非线性加权,减弱可见光图像特有的颜色信息,模拟红外光的红外辐射信息,实现输入级模态间差异的削减;接着模态迁移模块的输出和原始红外图像一起,通过共享参数的卷积神经网络来提取特征。文献[9]中已有实验论证,将ResNet50的最后一层下采样层

(down sampling layer)的池化步长(stride)从 2 修改为 1,可有效避免尺寸减小带来的信息损失,提高特征提取性能,因此在行人重识别的多项工作^[6-8]中都将其作为特征提取的基准分类模型。为了证明所提方法的有效性,同样选择该微改的 ResNet50(下面简称为 ResNet50)作为基础模型进行对比。另外,在去掉最后 ImageNet 的全连接层基础上,使用 ImageNet 大规模数据集上预训练的权重参数,进行迁移学习后所提取的特征经过多粒度特征分解模块,通过水平分块将特征分为两级细粒度特征。再将全局特征和细粒度特征分别进行三元组损失函数计算,经过全连接层进行交叉熵损失函数计算,完成整个网络的端到端的训练。

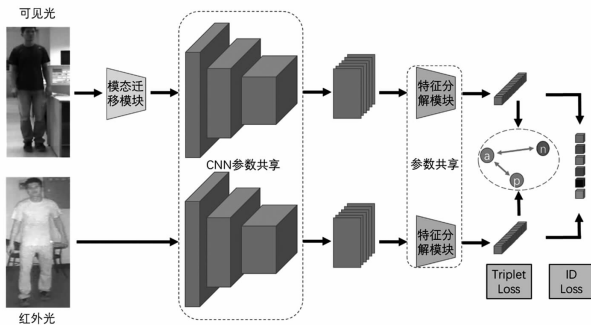


图 1 网络整体结构

2.2 模态迁移模块

由于不同模态图像灰度值反映的信息语义不同,且两种模态信息并没有已知的对应关系,因此,为了模拟行人图像的跨模态迁移减小图片级模态差异,使用注意力机制中的通道注意力,期望通过反向传播让模型拟合可见光模态到红外模态的通道级转换。具体结构如图 2 所示。

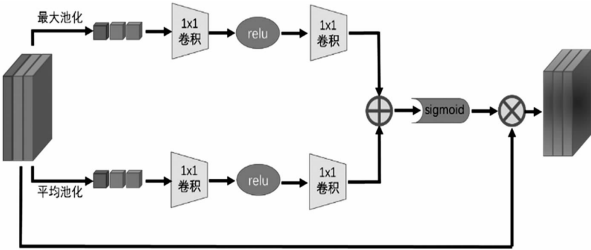


图 2 模态迁移模块结构

可见光图像首先经过最大池化和平均池化得到两个长度为 3 的一维向量特征表示,再经过 1×1 卷积和非线性的 relu 激活层再次叠加 1×1 卷积层用于拟合模态迁移变换。将两条路径得到的输入进行像素级叠加;再经过 sigmoid 激活函数提高非线性,得到不同通道的注意力权重;再与原始图像进行通

道级相乘,得到注意力加权后的三通道图像,用于后续特征提取。

2.3 特征分解模块

同一个行人在不同光照条件、视角差异、姿态变化等情况下摄像头所捕获到的图像有较大差异。若仅观察整体特征不利于行人身份的准确识别,如 2 个不同行人若身体姿态相似,则极易被误判为同一人,因此考虑行人图像的局部细粒度特征在一定程度上可提高模型鲁棒性。根据特征大小不同应当考虑不同粒度,如面部佩戴的眼镜、帽子等应该使用较细的粒度,裤子外观应该使用较粗粒度。据此设计了多粒度特征分解模块,如图 3 所示。前级特征提取得到了深层网络多维特征,分别经过 3 个不同分支,包括全局特征分支、粗粒度特征分支和细粒度特征分支。其中全局特征分支提取原始的图像整体特征,粗粒度特征分支将原特征进行水平分块为 3 组,细粒度特征分支将原特征水平分块为 6 组,总共得到 10 组特征,之后每组特征都经过全局平均池化层化为特征向量,再经过 1×1 卷积层进行降维,最后接批量归一化层(batch normalization, BN 层)将各组的特征实现通道级拼接得到一个总体多粒度特征,将其作为该输入图像的特征表示,用于后续损失函数计算和预测阶段的特征匹配。其中全局特征分支降维为 512 通道,粗粒度特征分支降维为 256 通道,细粒度特征分支降维为 128 通道。

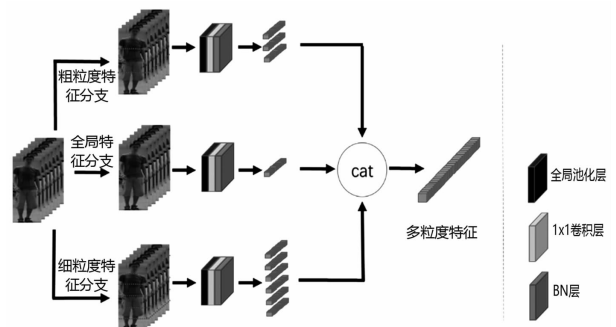


图 3 多粒度特征分解网络结构

3 实验结果与分析

3.1 数据集及评价指标

所选择的数据集为可见光-红外光跨模态行人重识别领域使用最广泛的公开数据集 SYSU-MM01,共包含 287 628 张可见光行人图片和 15 792 张红外光行人图像,来自于 6 个摄像头,包括室内场

景和室外场景。整个数据集共包含 491 个不同身份的行人,每个行人都至少有一张可见光图像和一张红外光图像。SYSU - MM01 数据集已经预先划分了训练集、测试集和验证集,其中训练集共 296 个 ID 的行人,测试集共 96 个,验证集共 99 个。目前普遍使用的处理方式是将原始的训练集和验证集合并到一起训练。测试集中的红外图像样本称为 query 集,也叫查询集,可见光图像样本称为 gallery 集,也叫待查集。测试模式包括全部场景(all)和室内(indoor)场景以及单发(single shot)和多发(multi-shot)两种情况互相组合。

评价指标使用累计匹配特性(cumulative matching characteristics, CMC)和平均精确度(mean average precision, mAP)两项指标,其中 mAP 指标在目标检测领域早已广泛使用,而 CMC 指标一般只在图像检索、重识别等问题中使用。模型在测试时将 query 集和 gallery 集全部样本输入网络提取各自的特征向量计算相似度,从而对于 query 集每个样本都能得到 gallery 集全部样本的相似度降序排列。对于 CMC 指标一般使用 rank - k 命中率进行计算。将 query 集中全部样本的 rank - k 计算平均值即可得到最终的 rank - k 结果。一般最常选择的 k 值为 1、10 和 20。

3.2 结果分析

实验采用的数据增强包括随机图像剪切(crop)、水平翻转和随机擦除,最后图像大小为 288 × 144 像素,网络使用 Adam 优化器,权重衰减系数为 5×10^{-4} ,初始学习率为 0.01,并且学习率衰减使用 warm-up 策略,即前 10 个 epoch(1 个 epoch 为将所有训练样本训练 1 次)学习率从 0.001 线性增长到 0.01,10 到 30 个 epoch 学习率保持 0.01,30 到 60 个 epoch 学习率为 0.001,60 到 80 个 epoch 学习率为 0.000 1。一共训练 80 个 epoch。ResNet50 由于使用了预训练参数,因此其学习率设置为模型学习率的 0.1 倍。使用 $P \times K$ 的采样策略,即每次迭代选择 P 个 ID,每个 ID 选 K 张可见光图像和 K 张红外光图像,实验设置 P 为 6, K 为 4。因此每次训练的 batch 大小为 48 张图片。模型在 SYSU - MM01 数据集的 4 种模式下进行测试,结果如表 1 所示。

根据表 1 可知,在全场景模式下,单发和多发两种情况下所提方法指标都高于现有方法,其中单发的 rank - 1 高于当前最佳方法 6.53%,mAP 也高出最佳方法 2.79%,多发模式下也有相同结果;并且室内场景下由于难度有所下降,各项指标都有所提高,但仍然优于现有其他方法:证明了所提方法的有效性。

3.3 对比实验

为进一步证明所设计的两种模块各自的有效

表 1 SYSU - MM01 数据集不同方法结果对比

方法	全场景								室内场景							
	单发				多发				单发				多发			
	r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP	r1	r10	r20	mAP
GSM	5.29	33.71	52.95	8.00	6.19	37.15	55.66	4.38	9.46	48.98	72.06	15.57	11.36	51.34	73.41	9.03
A - FC	9.30	43.26	60.38	10.82	13.06	52.11	69.52	6.68	14.59	57.94	78.68	20.33	20.09	69.37	85.80	13.04
Two - stream	11.65	47.99	65.50	12.85	16.33	58.35	74.46	8.03	15.60	61.18	81.02	21.49	22.49	72.22	88.61	13.92
One - stream	12.04	49.68	66.74	13.67	16.26	58.14	75.05	8.59	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Zero - padding	14.80	54.12	71.33	15.95	19.13	61.40	78.41	10.89	20.58	68.38	85.79	26.92	24.43	75.86	91.32	18.64
cmGAN	26.97	67.51	80.56	27.80	31.49	72.74	85.01	22.27	31.63	77.23	89.18	42.19	37.00	80.94	92.11	32.76
BDTR	27.32	66.96	81.07	27.32	—	—	—	—	31.92	77.18	89.28	41.86	—	—	—	—
eBDTR	27.82	67.34	81.34	28.42	—	—	—	—	32.46	77.42	89.62	42.46	—	—	—	—
D2RL	28.90	70.60	82.40	29.20	—	—	—	—	—	—	—	—	—	—	—	—
HPILN	41.36	84.78	94.51	42.95	47.56	88.13	95.98	36.08	45.77	91.82	98.46	56.52	53.05	93.71	98.93	47.48
XIV	49.92	89.79	95.96	50.73	—	—	—	—	—	—	—	—	—	—	—	—
所提方法	56.45	88.32	94.78	53.52	59.89	92.45	95.78	48.33	59.12	93.35	98.88	63.21	64.21	95.32	99.21	50.32

注:A - FC 为 Asymmetric - FC;“—”为原文未公布,加粗表示最佳效果。

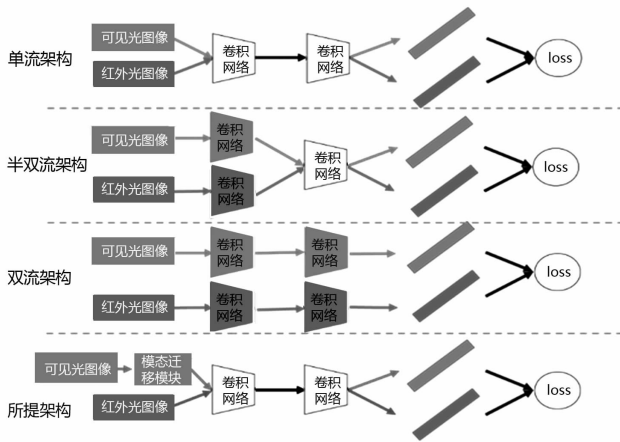


图 4 不同网络架构对比

性,对比实验设计包括不同网络架构的对比和不同基础模型的对比。对于跨模态行人重识别问题,基于深度学习的方法其模型可归类为 3 种网络架构:单流架构、双流架构、半双流架构,如图 4 所示。将 3 种网络架构和所提方法一起进行对比,为了控制变量,本部分实验都不采用所提出的特征分解模块,并且 4 种架构的基础模型均采用前面所述的 ResNet50,其中半双流架构的共享卷积层使用 ResNet50 的 layer1 - 4。在全场景单发模式下实验结果证明了使用所提的注意力模块后的网络架构的有效性,如表 2 所示。

表 2 不同网络架构对比

方法	r1	r10	r20	mAP
单流	40.85	81.13	88.23	37.45
双流	35.22	73.57	84.45	31.67
半双流	49.91	83.26	89.33	46.29
所提方法	52.14	85.3	92.56	47.83

为了对比所提出的特征分解模块的有效性,分别使用 4 种不同的经典预训练模型作为特征提取的基础模型,分别为 ResNet50、ResNet18、SqueezeNet、DenseNet121,每种基础模型都进行是否使用特征分解模块的对比。在全场景单发模式下实验结果如表 3 所示。

表 3 不同基础模型对比

方法	r1	r10	r20	mAP
ResNet18	50.22	83.35	89.89	48.27
DenseNet121	51.88	84.96	91.79	47.34
SqueezeNet	49.66	82.99	88.39	47.56
ResNet50	52.14	85.3	92.56	47.83
ResNet18'	53.88	86.56	93.23	51.21
DenseNet121'	55.74	87.68	93.99	51.77
SqueezeNet'	53.21	85.93	89.12	47.76
ResNet50'	56.45	88.3	94.78	53.52

注:“'”表示使用特征分解模块。

从表 3 结果可看出,不同的预训练模型效果有差别,但使用所提特征分解模块后各项指标均有所提高,证明所提特征分解模块的有效性。

4 结 语

针对可见光和红外光跨模态行人重识别问题,提出了一种新的网络结构,包括基于注意力的模态迁移模块用于缓解巨大的跨模态差异以及基于图像分块的多粒度特征分解模块用于提取图像全局和局部的多粒度特征。在 SYSU - MM01 公开数据集上进行的测试表明,所提两种模块能有效提高行人重识别模型精度,与现有多种方法比较达到了目前更好的效果。但与可见光单模态行人重识别相比,跨模态行人重识别模型效果仍然有很大的提升空间。考虑到深度学习中注意力机制有多种实现方式,接下来将进一步改进模态迁移模块,探究可见光和红外光之间更有效的特征变换结构,进一步减小跨模态差异,提高算法性能。

参考文献

- [1] Liang Zheng, Yi Yang, Alexander G Hauptmann. Person Re - identification: Past, Present and Future[J]. arXiv preprint arXiv:1610.02984, 2016.
- [2] A. Wu, W. S. Zheng, H. Yu, et al. RGB - infrared Cross - modality Person Re - identification[C]// in Proceedings of the IEEE International Conference on Computer Vision, 2017:5380 - 5389.
- [3] P. Dai, R. Ji, H. Wang, et a. Cross - modality Person Re - identification with Generative Adversarial Training [C]//in Proceedings of International Joint Conference on Artificial Intelligence, 2018: 677 - 683.
- [4] Mang Ye, Zheng Wang, Xiangyuan Lan, et al. Visible Thermal Person Re - identification via Dual - constrained Top - ranking[C]. In IJCAI'18:Proceedings of the 27th International Joint Conference on Artificial Intelligence, 2018:1092 - 1099.
- [5] Mang Ye, Xiangyuan Lan, Zheng Wang, et al. Bi - Directional Center - constrained Top - ranking for Visible Thermal Person Re - identification[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 407 - 419.

(下转第 15 页)

使用所提出的方法进行分类,结果如表2所示。

表2 变电站巡视周期分类结果

变电站编号	分类等级	变电站编号	分类等级	变电站编号	分类等级
S1	3	S13	3	S25	3
S2	4	S14	3	S26	3
S3	4	S15	3	S27	1
S4	4	S16	3	S28	4
S5	4	S17	2	S29	3
S6	1	S18	3	S30	3
S7	3	S19	3	S31	3
S8	2	S20	3	S32	2
S9	1	S21	4	S33	1
S10	2	S22	3	S34	2
S11	3	S23	3	S35	3
S12	3	S24	3		

由表2可看出,S6、S9、S27、S33共4座变电站被划分为第1类,S8、S10、S17、S32、S34共5座变电站被划分为第2类,S2、S3、S4、S5、S21、S28共6座变电站被划分为第4类,其他20座变电站被划分为第3类,验证了所提方法的有效性。分类结果完全取决于变电站本身实际情况,不受主观因素影响。

5 结语

针对差异化制定变电站集中监控全面监视周期问题,首先分析了变电站全面监视周期的影响因素,按影响作用由大到小包括变电站电压等级、集中监控缺陷发生频次、设备运行情况、变电站重要程度、变电站运行环境和设备平均负载6个方面。电压等级越高,集中监控缺陷发生频次越高,设备运行情况越差,

变电站重要程度越高,运行环境越恶劣,设备平均负载率越高,越应加强监视力度,缩短全面监视周期。

提出了基于竞争神经网络的变电站集中监控全面监视周期分类方法,其步骤可概括为:将变电站全面监视周期的6个影响因素进行指标量化,得到表征每座变电站全面监视周期影响因素的六维特征向量;建立竞争神经网络,输入层神经元个数为6个,输出层神经元个数取变电站全面监视周期想要分类的类别数;使用Matlab工具箱函数competlayer实现竞争神经网络分类;根据竞争神经网络给出的分类结果,确定每座变电站的全面监视周期。

所提方法可利用竞争神经网络实现变电站集中监控全面监视周期的科学合理分类,为供电公司差异化安排变电站集中监控全面监视周期提供技术支持。

参考文献

[1] 谢金泉,李晓华,何毅,等. 变电设备状态巡视策略研究[J]. 广东电力, 2012, 25(9):5-9.

[2] 陈明. Matlab神经网络原理与实例精解[M]. 北京:清华大学出版社,2013.

[3] 董长虹. Matlab神经网络与应用[M]. 北京:国防工业出版社,2007.

[4] 陈宏彬,李悦宁,刘琨. 基于状态检修的无人值班变电站状态巡视管理[J]. 现代营销(经营版), 2019(4):150.

[5] 王少博,刘辉,张兰钦,等. 基于设备状态分析的无人值班变电站巡视模式研究[J]. 河北电力技术, 2013,32(2):10-12.

[6] 彭业. 引入状态评估分析的变电站设备巡视优化策略的研究[D]. 广州:华南理工大学,2012.

作者简介:

张宇泽(1990),男,硕士,工程师,研究方向为电力系统分析与控制。
(收稿日期:2020-07-06)

(上接第10页)

[6] J. W. Lin, H. Li. HPILN: A Feature Learning Framework for Cross-modality Person Re-identification[J]. arXiv preprint arXiv:1906.03142, 2019.

[7] Z. Wang, Y. Zheng, Y. Y. Chuang, et al. Learning to Reduce Dual-level Discrepancy for Infrared-visible Person Re-identification [C]//in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 618-626.

[8] Diangang Li, Xing Wei, Xiaopeng Hong, et al. Infrared

-visible Cross-modal Person Re-identification with An X Modality[C]// The Thirty-Fourth AAAI Conference on Artificial Intelligence (AAAI-20), 2020.

[9] Hao Luo, Wei Jiang, Youzhi Gu, et al. A Strong Baseline and Batch Norm Neutralization Neck for Deep Person Reidentification[J]. IEEE Transactions on Multimedia, 2020,22(10):2597-2609.

作者简介:

张森(1995),男,硕士研究生,研究方向为计算机视觉。
(收稿日期:2020-10-26)