UDC 004.932

人员的陪伴和重新识别 在智能视频监控系统中 使用卷积神经网络

R. P.博古什1-, S.A.伊格纳季耶娃1个, S. V. Ablameyko2.3
1个波洛茨克国立大学 波洛茨克的 Euphrosyne，
白俄罗斯新波洛茨克
r.bogush@psu.by ;
2个白俄罗斯国立大学，明斯克；
3个白俄罗斯国家科学院信息学问题联合研究所，明斯克

介绍。目前，视频监控系统的使用有所增加。 观察，这可以通过他们解决的广泛任务以及为此不断开发的算法和硬件支持来解释。值得注意的是，由于硬件基础的快速提升，摄像机分辨率的提高，通信信道带宽的增加，5G技术的引入，人工智能处理信息的方法的发展和应用、处理大量数据的技术、云解决方案、物联网信息转网，这种趋势在未来还会继续。

在视频监控系统中，最有效的是基于使用空间分离的 IP 摄像机和多代理架构的空间分布式系统。此类系统使用视频数据分析，其智能在于能够自动分析视频流以识别特定对象或其动作。在这些任务中，多人对一台摄像机形成的视频序列的跟踪及其重新识别是重要且相关的 [1]。人的重识别（re-identification, inter-camera tracking）可以定义为对从空间分离的相机获得的同一个人的所有图像分配相同的名称或索引的任务，这些图像的可见区域不与彼此，基于其图像特征的选择和分析。在空间分布的视频监控系统中使用重新识别允许收集关于一个人在由多个视频监控摄像机覆盖的大面积上的独特出现次数的统计数据。基于人的跟踪和重新识别，可以实现各种实际任务：在“智能家居”和“智能城市”系统中监控人和其他物体的移动，在自动驾驶系统中分析环境，医学和运动中运动正确性的评估、工业视觉系统中的对象跟踪、监控和安全系统中人类活动类型的识别等。在空间分布的视频监控系统中使用重新识别允许收集关于一个人在由多个视频监控摄像机覆盖的大面积上的独特出现次数的统计数据。基于人的跟踪和重新识别，可以实现各种实际任务：在“智能家居”和“智能城市”系统中监控人和其他物体的移动，在自动驾驶系统中分析环境，医学和运动中运动正确性的评估、工业视觉系统中的对象跟踪、监控和安全系统中人类活动类型的识别等。在空间分布的视频监控系统中使用重新识别允许收集关于一个人在由多个视频监控摄像机覆盖的大面积上的独特出现次数的统计数据。基于人的跟踪和重新识别，可以实现各种实际任务：在“智能家居”和“智能城市”系统中监控人和其他物体的移动，在自动驾驶系统中分析环境，医学和运动中运动正确性的评估、工业视觉系统中的对象跟踪、监控和安全系统中人类活动类型的识别等。

此类任务的特点是实现复杂性高，需要准确定位帧中的人，并正确识别当前帧或另一台摄像机相对于先前帧的帧。主要问题之一是描述一个人的描述符的选择 [2]。为了解决这个问题，有必要识别不同的特征，并通过比较来自不同帧的人的图像或通过执行查询，将它们相互比较或与现有的许多人图像样本中的特征（用于重新识别的画廊）进行比较). 图像中对象（包括人）的一组最显着特征的搜索和选择不是形式化的。因此，需要进行经验搜索

标志，这在大多数情况下是一个漫长而费力的过程。为了陪伴和重新识别人，由于不同角度的外观模糊、光照变化、不同的相机分辨率、遮挡，这种方法需要不合理的大量时间。因此，长期以来，这些问题都没有取得明显的成果。计算机技术的改进和深度学习领域的发现，特别是卷积神经网络 (CNN) 的发展，使得提取人物图像特征的过程自动化成为可能，并显着提高了重新识别的准确性。 -鉴定，不过，目前还没有完全得到解决方案。

跟踪和重识别的原理和问题。临- 空间分布式视频监控系统由地理上分散的 IP 摄像机组成，通常基于单个数据处理中心进行组织。图 1 显示了具有人员跟踪和重新识别功能的智能视频系统的总图。

在每一帧F_q从$和_{1个}$,$和_{2个}$， C_q网络摄像机，q- 系统中的摄像机编号，s
在基于 SNS 的检测器的帮助下，检测摄像机视野内的所有人，并为他们形成边界框，用矩形描述检测到的图形。对于一个人的每一张照片

$我_{我}$， 在哪里$我$=1,...,$否_{图片}$.$否_{图片}$- 使用另一个 SNA 的图像总数 确定 CNN 特征的向量$F_{基因}$
（SNA 描述符）构成一般
$_{我}$
SNS 特征空间 =-$F_{基因}$ $_{我我}$ $_{我}$ -,$我$=1,...,$否_{图片}$. 这组描写
ditch以表格的形式呈现，其中每一行都是一个SNA描述符$F_{基因}$对于一张图片。

$_{我}$
对于视频序列中移动的人，可以更改一个或多个参数：帧上的坐标$\left(\chi_{Fk}\right.$
$_{F_{基因}}$, $是_{F_{基因}}$$\left.\right)$, 尺码$\left(深圳_{Fk}\right)$,$_{F_{基因}}$形式
$\left(FR_{F_{基因}}\right)$ 在一定的时间间隔$(吨)$. 其形式的改造和（或） 尺寸改变了它在框架上的特征（$F_{基因}$). 因此，移动物体
$_{我}$
被形容为：

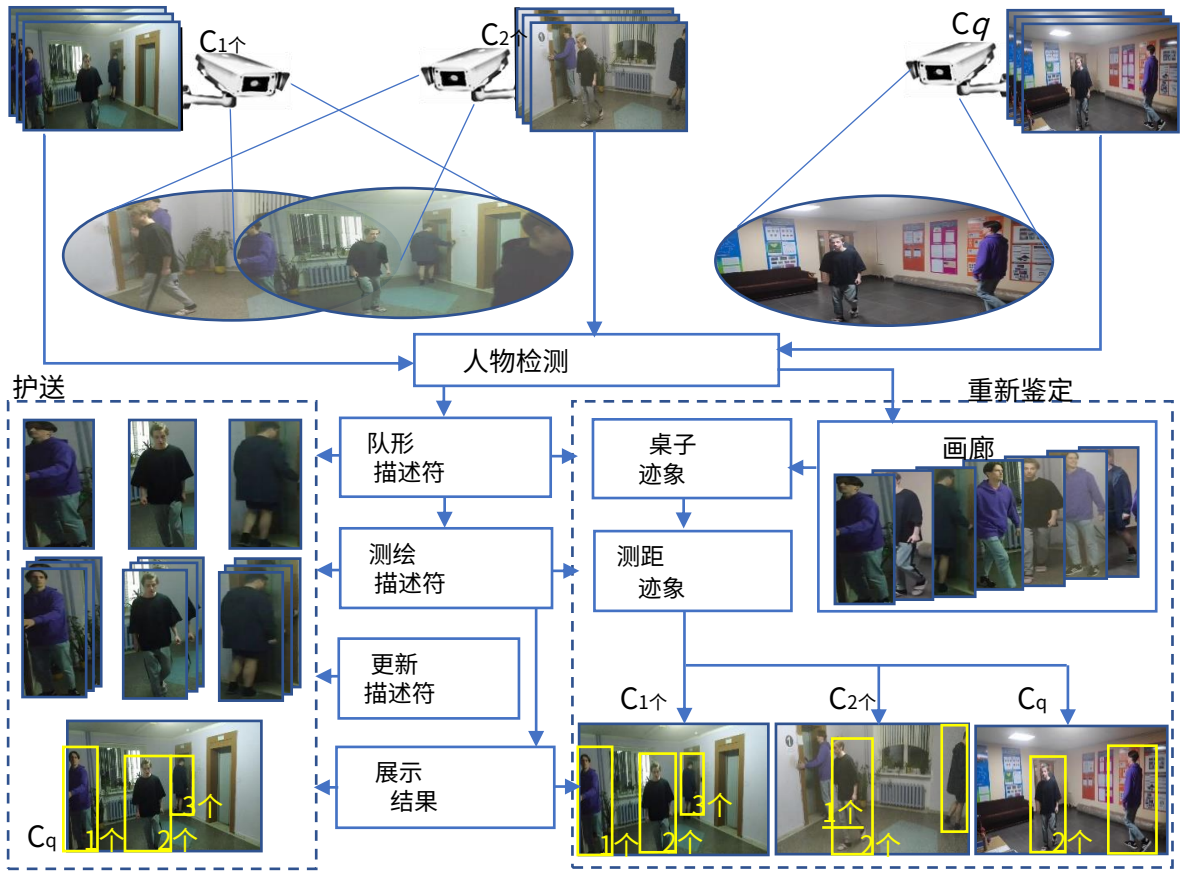
$$欧巴_{丁}=-F_{基因},X_{Fk}{}_{F_{基因}}, 是_{F_{基因}}, Ns_{F_{基因}}-.$$

在一个人的陪同下，可以理解他在一台摄像机形成的视频序列的每一帧上的位置，在时间间隔$(吨)$.

需要注意的是，包括人在内的物体跟踪有两类不同复杂度的任务：跟踪一个物体（Visual object tracking，VOT）和跟踪多个物体（Multiple object tracking，MOT）。第一种情况的特点是，一个给定的对象，一个人，被检测到并定位在某个帧上，而其他人没有被呈现为感兴趣的对象，也没有被检测到。

在画面中有多个伴奏的情况下，通常会有几个人同时移动或静止一段时间。而且，他们中的许多人可能有视觉上相似的标志，短时间离开现场，或者完全离开现场，而其他人实际上可以出现在先前的出口处，例如在门口房间的入口。-schenie。因此，由于人员彼此交叉或他们躲在背景元素后面，护送中断的可能性很高。因此，这种实时跟踪是一项非常困难的任务。

求解时，在检测到帧上的人物图形后，在帧的空间区域和视频序列上的时间区域计算分析所选片段的特征。这些可能包括：SNS 标志、直方图、颜色标志；框架中人物选定区域的中心坐标；当前帧相对于前一帧的移动方向；前一帧区域的宽度和高度；运动轨迹；旅行时间。可以为一个人的整个图像和（或）它的各个部分计算类似的特征。对于所有跟踪对象和在当前帧上找到的对象，计算相似度值，在此基础上建立检测对象和跟踪对象之间的对应关系。通过解决将检测到的区域分配给现有跟踪对象的问题来确定检测到的人及其在先前帧上的位置的对应关系。为此，在检测器检测到的区域与现有跟踪对象之间形成一个相似度矩阵。作为输出数据，形成一个向量，其中每个对象都被分配了检测器检测到的对象的索引。轨迹是在第一次检测到一个人时创建的，如果在一定数量的连续帧中未检测到此人并且没有与他的先前帧进行比较，则删除该轨迹，即 据信他离开了由摄像机拍摄的舞台。为此，在检测器检测到的区域与现有跟踪对象之间形成一个相似度矩阵。作为输出数据，形成一个向量，其中每个对象都被分配了检测器检测到的对象的索引。轨迹是在第一次检测到一个人时创建的，如果在一定数量的连续帧中未检测到此人并且没有与他的先前帧进行比较，则删除该轨迹，即 据信他离开了由摄像机拍摄的舞台。为此，在检测器检测到的区域与现有跟踪对象之间形成一个相似度矩阵。作为输出数据，形成一个向量，其中每个对象都被分配了检测器检测到的对象的索引。轨迹是在第一次检测到一个人时创建的，如果在一定数量的连续帧中未检测到此人并且没有与他的先前帧进行比较，则删除该轨迹，即 据信他离开了由摄像机拍摄的舞台。如果给定的人在一定数量的连续帧中没有被检测到并且没有与他之前的帧进行比较，则被删除，即 据信他离开了由摄像机拍摄的舞台。如果给定的人在一定数量的连续帧中没有被检测到并且没有与他之前的帧进行比较，则被删除，即 据信他离开了由摄像机拍摄的舞台。



米。一、智能伴奏视频系统总体方案
和重新识别人

为了在重新识别过程中描述一个人，描述符可以表示为：

$$P_{ID}\left(p_{n}^{ID}, F_{我}^{基因}, F_{我}^{添加}\right),$$

在哪里p_{n}^{ID} – 一个人的标识符（标签）；n是可能的标识符的数量
ditch 等于独特的人的总数；$F_{我}^{基因}$– SNA 标志*我*日
一个人的形象可以分为全局的，将他的形象作为一个整体来表征，以及局部的，这是通过将图像分成部分来获得的;
$F_{我}^{添加}$- 可能包含允许的信息的附加标志 提高重新识别系统的效率，例如，标识符

相机C_{ID}, 帧数Cq第摄像机F_q， 时间*吨*$_{Fq}$ $_{米}$ $_{米}$接收帧*米*Cq日
摄像机。

为了重新识别的实际实施，创建了一个包含人物图像及其描述符的表，称为图库。在收到一个人的重新识别请求后，计算他的特征向量，用于找到距离d_q, 决定相似度

在给定的请求和图库图像描述符之间。使用找到的距离，在表中进行排名$d_{分钟}$前$d_{最大限度}$. 考虑在内
额外的特征，图像被排除在外，根据一些标准，允许我们假设，尽管视觉特征相似，但候选图像不是所需的图像。从特征表中排除所有不合适的候选者后，重新识别的结果显示人物图像。$F_{基因}$在排名表列表的顶部。

$_{我}$
排名列表中的第一个人作为重新识别的结果，与查询最相似。

提高跟踪和重识别的准确性. 众所周知，基于 SNS 形成的人类描述符的有效性取决于其体系结构和执行训练的数据集。增加 SNS 层数可以提高工作的准确性。

在比较人时，需要考虑他们相似和不同特征的可变性，并确保可接受的计算成本。

表 1. CNN 架构的比较

社交网络类型	*数量 图层*	*可能性 遇到的错误 顶1*	*可能性 中的错误 top5 指标*	*速度（毫秒）*
亚历克斯网	8个	42.90	19.80	14.56
盗梦空间-V1	22	-	10.07	39.14
VGG-16	16	27.00	8.80	128.62
VGG-19	19	27.30	9.00	147.32
ResNet-18	18	30.43	10.76	31.54
ResNet-34	34	26.73	8.74	51.59
ResNet-50	50	24.01	7.02	103.58
ResNet-101	101	22.44	6.21	156.44
ResNet-152	152	22.16	6.16	217.91
ResNet-200	200	21.66	5.79	296.51

根据[3]的测试结果，包含执行测试的计算机的特征，基于表1的分析，为了计算描述符和确保实时操作，ResNet-34 CNN是令人感兴趣的，层数少，计算精度令人满意。ResNet-34 中关闭连接的存在允许您更改层数以获得更好的训练结果。但是，考虑到支持的具体情况，需要选择不同人的特征进行后续比较，考虑到他们属于同一类“人”，这将不允许 ResNet-34 做所以。在 [4] 中，提出了一种改进的 CNN 架构：移除过滤器大小为 [7×7] 的输入卷积层，由于使用最小尺寸 [3×3] 的卷积核可以在重新识别中获得最佳结果 [153]；最终全连接层的输出个数减少到128个，这样就可以形成相同数量的特征来描述一个人；将 CNN 卷积层的数量减少到 29 个，核心尺寸为 [3×3]，在每层之后使用闭合连接。使用这种架构可以提高室内视频监控期间跟踪人员的准确性 [4]。

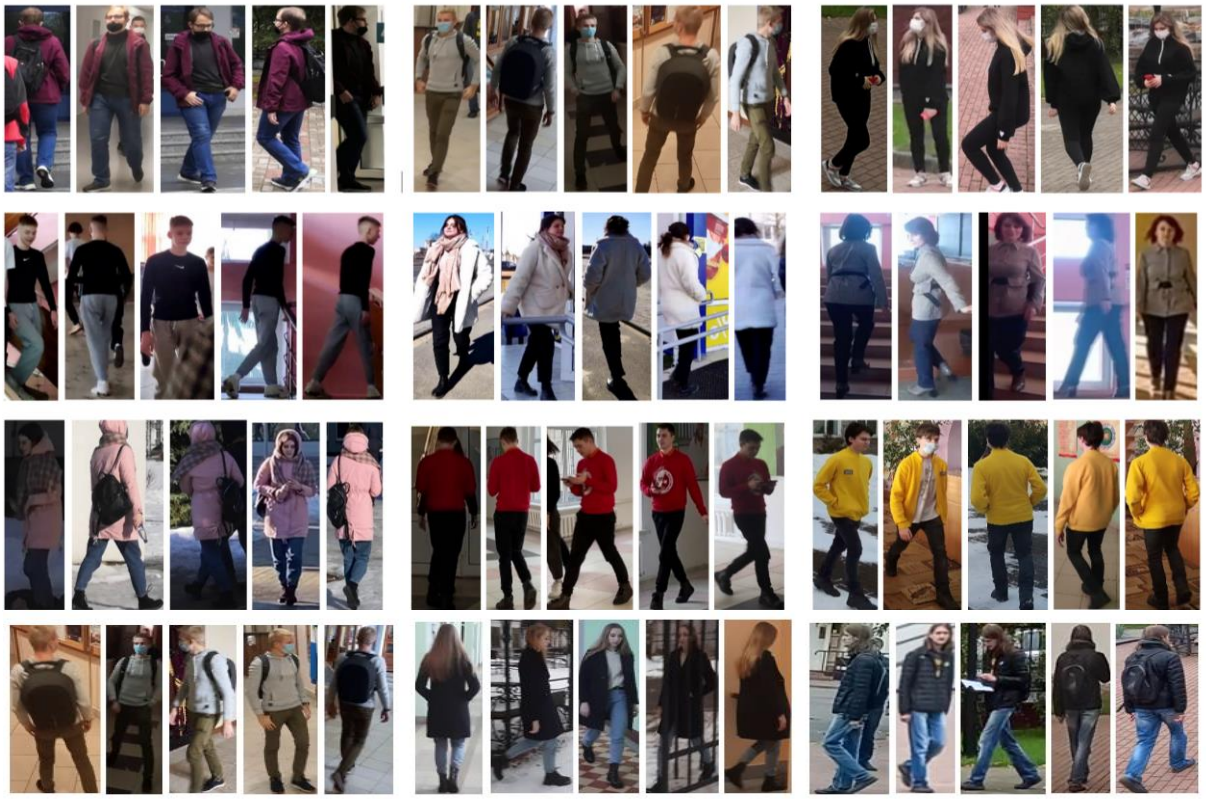
对于深度 CNN，在训练期间，可能会出现梯度爆炸或梯度消失等现象。它们导致在累积较大误差梯度时出现的问题，由于CNN权重更新非常快，因此网络模型不稳定。另一种类型的梯度消失会导致一个逆向问题，在该问题中也无法进行有效的学习。解决这些问题有不同的方法，其中之一是搜索 FA 激活函数。为特定应用任务选择 FA 涉及实验研究，这些研究将确定在准确性和时间成本方面最有效的方法。因此，分析最常见的FAs ReLU, Leaky-ReLU, PReLU, RReLU, ELU, SELU, GELU, Swish, Mish, 在卷积神经网络中用于人的重新识别，这是针对三个 CNN ResNet-50、DenseNet-121 和 DarkNet-53 进行的。对获得的结果进行分析后发现，激活函数 ReLU 和 GeLU 最有希望用于重复识别 [5]。然而，ReLU 的运行速度和结果的可重复性高于 GeLU [5]。

对数据集的主要要求是大量的图像，它们的多样性和均匀性。数据不足会导致过度拟合，即 CNN 训练示例的记忆和模型对新数据的不稳定性。数据集的多样性意味着图像是从具有不同特性（分辨率、视角、安装位置）、不同视频监控条件（一天中的时间、季节、照明）和不同外貌（性别、身高）的人获得的。 ., 体质, 衣服). 多样性的增加增加了提取特征的可靠性和训练系统对不熟悉数据的稳定性。均匀性意味着应该有大约相等数量的各种例子，因为 大量具有相似特征的图像会导致泛化不平衡，即 系统将认为为相似图像选择的特征比为不足够的示例选择的特征更重要。

对于 SNS 训练，形成了 PolReID 数据集 [6]，其中每个人使用位于不同位置的 2 到 10 个摄像头，以及每个摄像头在不同天气条件和季节（夏季，秋季和冬季），室内有不同强度的自然和人工照明

强度。要从帧中提取边界框，请使用 YOLOv4 检测算法。视觉分析后，操作员删除了不正确的边界框。对于数据集中的每个人，都有部分水平和垂直重叠的图像。从不同的角度呈现同一个人。数据集总共包含 657 个人的图像，包括 52035 张图像。

PolReID 分为训练数据和测试数据。对于训练，使用了 398 个不同人的 ID（32516 个边界框），用于测试 - 259 个 ID（19519 个边界框）。PolrReID 包括 440 名男性和 217 名女性的图像；18至30岁的524人和30岁以上的133人。340 人的图像是在室内拍摄的，214 人的图像来自室外摄像机，103 人的图像来自室内和室外摄像机。210 人戴着口罩，其中 33 人被一些没有戴口罩的摄像机记录下来。夏季拍摄了 95 人，冬季拍摄了 288 人，春季和秋季拍摄了 274 人。图像示例如图 2 所示。



米。2. 来自 PolReID 数据集的样本图像

接下来，合并现有集 Market[7]、Duke[8]、CUHK02[9]、CUHK03[10]、MSMT17[11] 和 PolReID。在生成的大型数据集上，对 DenseNet-121、ResNet-50、PCP CNN 进行了训练，并使用 Rank1 和 mAP 指标评估了准确性，结果如表 2 所示。对表 2 的分析表明，创建的数据集改进了所有测试的重新识别指标，获得了 PolReID Rank1 = 95.41，mAP = 84.74 的最大值。SNA PCB 在源域和目标域匹配时最有效，对于 PolReID 和 Market1501 进行跨域重识别也是如此。DenseNet-121 在 DukeMTMC-ReID 上表现最好，在 Market-1501 和 DukeMTMC-ReID 上训练时在池化数据集上表现最好。

表 2. 实验结果

数据为了测试	*国民账户体系*	*训练数据*							
		市场-1501		*杜克MTMC-ReID*		*MSMT17*		*联合数据集*	
		评估准确性的指标							
		Rank1	*地图*	*Rank1*	*地图*	*Rank1*	*地图*	*Rank1*	*地图*
市场1501	稠密-网121	88.86	73.01	49.23	21.71	54.22	26.40	94.09	83.34
	ResNet-50	83.33	71.16	43.88	18.68	48.49	22.81	92.12	80.62
	电路板	92.70	77.69	55.05	25.89	55.53	25.74	93.14	81.62
公爵MTMC-里德	稠密-网121	37.21	20.18	81.51	64.81	55.61	34.51	86.45	74.00
	ResNet-50	30.57	15.86	79.04	62.40	50.76	30.84	84.20	71.19
	电路板	40.44	22.23	84.87	70.30	54.35	33.26	86.36	73.86
MSMT17	稠密-网121	12.72	03.92	19.84	5.94	70.53	40.99	76.73	51.13
	ResNet-50	9.24	2.68	15.04	4.32	65.71	36.56	72.05	45.64
	电路板	11月06日	3.10	16.49	4.57	70.42	42.81	73.87	48.17
极地ID	稠密-网121	63.66	34.55	74.21	43.44	83.64	58.09	95.25	83.82
	ResNet-50	57.61	29.39	67.85	37.16	79.69	52.91	94.12	80.89
	电路板	62.61	35.31	72.20	40.80	86.38	60.62	95.41	84.74

结论。使用 SNS 形成多发性人的体征 支持和重新识别使得在多摄像机视频监控系统中实际执行这些任务成为可能。由于新的 CNN 架构和用于训练的大型复合数据集，本文考虑的方法旨在提高其准确性。所呈现的研究结果表明，可以改进对人员的跟踪和重新识别。

使用的来源列表

1. Behera，NKS 智慧城市行人再识别：最新技术和
前方道路/NKS Behera、P. Kumar、S. Bakshi // 模式识别快报。- 2020. - 卷。138. – 第 282-289 页。

2. Ye, S. Person Tracking and Re-Identification in Video for Indoor Multi-Camera Surveillance Systems / S. Ye、R. Bohush、C. Chen、I. Zakharava、S. Ablameyko // 模式识别和图像分析，2020 年。- 卷。30，第 4 期 - 第 827-837 页

3. 流行 CNN 模型的基准 [电子资源] – 2020 – 交流模式
cess：https://github.com/jcjohnson/cnn-benchmarks。– 访问日期：09/16/2022。

4. Bohush, R. Robust Person Tracking Algorithm Based on Convolutional Neural Net-室内视频监控工作 / R. Bohush, I. Zakharava// 计算机和信息科学通信。- 2019. - 卷。1055.-P. 289–300

5. Chen, H. 重复条件下卷积神经网络激活函数的选择
视频监控系统中的人员识别 / S. Ignatieva, R. Bogush, S. Ablameiko // 编程。- 2022 年 - 第 5 号 - 第 15-26 节。

6. PolReID [电子资源] - 2022 - 获取方式：https://github.com/
SvetlanaIgn/PolReID.– 访问日期：2022 年 9 月 16 日。

7. 可扩展行人再识别：基准/L. Zheng [等。al] // IEEE Proc
计算机视觉国际会议 (ICCV)。- 2015 年 - 第 1116-1124 页。

8. 多目标、多摄像机跟踪的性能测量和数据集 [Elec-
电子资源]。– 访问方式：https://arxiv.org/abs/1609.01775。– 访问日期：09/03/2022。

9. Li W, Wang X. 跨视图的局部对齐特征变换 / W. Li, X. Wang
//过程。IEEE 计算机视觉和模式识别会议。- 2013 年 - 第 3594-3601 页。

10. DeepReID：用于行人再识别的深度过滤配对神经网络 /
W. Li [等] // Proc。IEEE 计算机视觉和模式识别会议。- 2014 年 - 第 152-159 页。

11. Wei L, Zhang S, Gao W, Tian Q. Person Transfer GAN to Bridge Domain Gap for
行人再识别 / L. Wei [et al.]// Proc. IEEE/CVF 计算机视觉和模式识别会议。- 2018 年。- 第
79-88 页。