中图法分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-25

论文引用格式:Yang Hangyuan, Zhang Yongfei, Chen Siyu, Ma Yuzhuo, Zhang Yujia, Li Shangze, Bai Binyi, Hao Xiaohui, Bakana Sibusiso Reuben. Governed prototype knowledge transfer for exemplar-free lifelong person re-identification[J/OL]. Journal of Image and Graphics, XXXX:1-25. DOI: 10.11834/jig.260402. (杨航远, 张永飞, 陈思宇, 马愈卓, 张雨佳, 李尚泽, 白彬沂, 郝晓辉, 罗本. 受治理原型知识迁移的无样本终身行人再识别[J/OL]. 中国图象图形学报,XXXX:1-25. DOI: 10.11834/jig.260402.)[DOI:10.11834/jig.260402]

受治理原型知识迁移的无样本终身行人再识别

杨航远[1],张永飞[1,2*],陈思宇[3],马愈卓[1],张雨佳[1],李尚泽[1],白彬沂[1],郝晓辉[1],罗本[4]

1. 北京航空航天大学计算机学院,北京　100191; 2. 北京航空航天大学虚拟现实技术与系统全国重点实验室,北京　100191; 3. 真空电子技术国家重点实验室,北京　100015; 4. 南非科学与工业研究理事会,茨瓦内　0001

摘　要:**目的**　现有基于原型的无样本终身行人再识别(lifelong person re-identification,LReID)方法通常静态、等权复用历史原型,忽视领域相关性、统计可靠性和边界风险差异,易引入噪声迁移并加剧灾难性遗忘。为此,本文提出受治理原型知识迁移(governed prototype knowledge transfer,GPKT)框架,对历史原型进行选择、校准和可信度加权。**方法**　在不保存历史图像的条件下,GPKT以身份级分布原型、领域级描述符和样本支持统计保存历史知识;按领域与身份相关性检索原型并通过残差矩匹配缓解特征空间漂移,结合样本支持、方差稳定性、领域相关性和边界安全性生成治理权重,再将其注入原型诱导的关系蒸馏以迁移历史身份关系结构。**结果**　在12个行人再识别数据集(5个已见域、7个未见域)和两种训练顺序上,将GPKT与9种近期代表性方法进行比较。相较于相同50层残差网络(50-layer residual network,ResNet-50)骨干下综合平均性能最强的分布感知知识对齐与原型化(distribution-aware knowledge aligning and prototyping,DKP++)方法,GPKT在顺序-1/顺序-2的已见域平均的平均精度均值(mean average precision,mAP)/Rank-1(R1)准确率分别提升3.1/2.7和3.0/2.9个百分点,未见域平均分别提升3.2/2.6和3.3/2.6个百分点。将骨干替换为视觉Transformer(vision Transformer-base/16,ViT-B/16)后,顺序-1/顺序-2的已见域平均mAP分别较ResNet-50版本进一步提升1.5/1.5个百分点,未见域平均mAP分别提升1.8/1.9个百分点。逐阶段评估表明,GPKT在两种顺序下的平均遗忘率较DKP++降低约2个百分点,其最终性能优势由旧领域保持与新领域适应共同贡献。**结论**　GPKT通过相关性、可靠性与风险感知的原型治理和关系蒸馏,在不存储历史样本的前提下有效兼顾旧知识保持、新领域适应与跨域泛化。

关键词:灾难性遗忘;无样本终身学习;行人再识别;原型治理;原型记忆

Governed prototype knowledge transfer for exemplar-free lifelong person re-identification

Yang Hangyuan[1], Zhang Yongfei[1,2*], Chen Siyu[3], Ma Yuzhuo[1], Zhang Yujia[1], Li Shangze[1], Bai Binyi[1], Hao Xiaohui[1], Bakana Sibusiso Reuben[4]

1. School of Computer Science and Engineering, Beihang University, Beijing 100191, China; 2. State Key Laboratory of Virtual Reality Technology and Systems, Beihang University, Beijing 100191, China; 3. National Key Laboratory of Science and Technology on Vacuum Electronics, Beijing 100015, China; 4. Council for Scientific and Industrial Research, Tshwane 0001, South Africa

Abstract: **Objective** Person re-identification (ReID) aims to retrieve the same pedestrian across non-overlapping cameras

收稿日期:2026-07-03;**修回日期**:2026-10-22

*通信作者:张永飞 yfzhang@buaa.edu.cn

基金项目:国家自然科学基金(编号:62472016);中央高校基本科研业务费专项资金。

Supported by: National Natural Science Foundation of China (No. 62472016); the Fundamental Research Funds for the Central Universities.

© 中国图象图形学报版权所有

and is a key technique for intelligent video surveillance, smart cities, and public security. In real deployments, however, cameras, illumination, backgrounds, and even the identities to be recognized change continually, so a ReID model must keep learning from sequentially arriving domains without forgetting the knowledge it has already acquired. Lifelong person re-identification (LReID) is formulated for this purpose, and its central difficulty is catastrophic forgetting. Because the long-term storage of pedestrian images conflicts with privacy protection, storage cost, and regulatory compliance, exemplar-free LReID, in which no historical raw image can be accessed after a training stage ends, has become an important and practical research direction. Existing prototype-based exemplar-free methods compress historical identities into prototypes and reuse them to preserve old decision boundaries or to guide current feature learning. However, they usually treat all historical prototypes as static and equally trustworthy knowledge sources. As learning stages accumulate, historical prototypes differ markedly in domain relevance, statistical reliability, and boundary risk, and feature-space drift further misaligns the prototypes extracted by earlier models with the current representation space. Transferring these prototypes uniformly injects irrelevant, noisy, and boundary-ambiguous supervision into current training, which weakens old-domain retention and disturbs new-domain adaptation at the same time. We therefore argue that historical prototypes should not only be constructed but also governed, and we propose a governed prototype knowledge transfer (GPKT) framework that selects, calibrates, and trust-weights historical prototypes before they are allowed to influence the current model. **Method** GPKT follows a "select-calibrate-govern-distill" pipeline built on a standard ReID backbone. First, a hierarchical distribution prototype memory is constructed at the end of each stage without storing any historical image. It records an identity-level distribution prototype, namely the feature mean and per-dimension variance, for every historical identity, a domain-level descriptor consisting of the global feature mean and variance for every historical domain, and sample-support counts that indicate how reliably each prototype is estimated. Second, relevance-guided prototype selection and calibration performs coarse-to-fine retrieval. The most relevant historical domains are retrieved according to the similarity of feature centers and the consistency of log-variances between each domain descriptor and the current mini-batch, and the identity prototypes closest to the current identity centers are then kept within the selected domains. The selected prototypes are calibrated by residual moment matching, which preserves the relative position of each prototype inside its source domain while shifting and rescaling it to the current mini-batch statistics; a midpoint update and variance clipping further prevent over-correction caused by noisy batch statistics. Third, reliability- and risk-aware prototype governance assigns a continuous governance weight to each calibrated prototype. The reliability score increases with sample support and penalizes abnormal variance, the boundary-safety score combines a proximity gate and a margin gate to suppress prototypes that lie close to several current identity centers, and a softmax-normalized domain-level prior controls the overall contribution of each selected domain. Finally, governed relation distillation injects the governance weights as a logarithmic prior into the sample-to-prototype assignment, builds a prototype-induced relation distribution among current samples as the teacher, and aligns the relation distribution of current features with this teacher through the Kullback-Leibler (KL) divergence. Because relation structure, rather than identity-level alignment, is transferred, no explicit identity correspondence across domains is required. All prototype-related operations are performed only during training, so GPKT remains a standard ReID feature extractor at inference time. **Result** Experiments are conducted on the standard LReID benchmark, which contains 12 person ReID datasets, including 5 seen domains for lifelong training and 7 unseen domains for generalization evaluation, under two training orders, and GPKT is compared with 9 recent representative exemplar-free methods. With the same 50-layer residual network (ResNet-50) backbone, GPKT achieves the best seen-domain and unseen-domain average performance under both orders. Compared with distribution-aware knowledge aligning and prototyping (DKP++), the strongest competitor on average, GPKT improves the seen-domain average mean average precision (mAP)/Rank-1 accuracy by 3.1/2.7 and 3.0/2.9 percentage points (pp), and improves the unseen-domain average mAP/Rank-1 (R1) accuracy by 3.2/2.6 and 3.3/2.6 pp under Order-1 and Order-2, respectively. Replacing ResNet-50 with a base-size vision Transformer using 16×16 patches (ViT-B/16) further raises the seen-domain average mAP by 1.5/1.5 pp and the unseen-domain average mAP by 1.8/1.9 pp. Stage-wise evaluation shows that GPKT reduces average forgetting by about 2 pp relative to DKP++ and raises learning accuracy by 1.4 to 1.5 pp; the decomposition of the final performance further indicates that the final gain stems jointly from old-domain retention and new-domain adaptation, with retention contributing slightly more. A full-factorial ablation

© 中国图象图形学报版权所有

verifies that prototype selection, residual calibration, and governed distillation are complementary, and the complexity analysis shows that GPKT introduces no extra inference overhead, while its additional training computation is bounded by the fixed numbers of selected domains and prototypes and its prototype memory occupies only 137.5 mebibytes (MB) after five stages. **Conclusion** By combining relevance-, reliability-, and risk-aware prototype governance with relation distillation, GPKT turns the use of historical prototypes from indiscriminate reuse into selective, calibrated, and trust-weighted knowledge transfer. Without storing any historical exemplar, it effectively balances old-knowledge retention, new-domain adaptation, and cross-domain generalization, providing a practical solution for privacy-constrained lifelong person re-identification.

Key words: catastrophic forgetting; exemplar-free lifelong learning; person re-identification (ReID); prototype governance; prototype memory

论文引用格式: Yang Hangyuan, Zhang Yongfei, Chen Siyu, Ma Yuzhuo, Zhang Yujia, Li Shangze, Bai Binyi, Hao Xiaohui, Sibusiso Reuben Bakana. Governed prototype knowledge transfer for exemplar-free lifelong person re-identification[J/OL]. Journal of Image and Graphics. DOI: 10.11834/jig.260402. (杨航远,张永飞,陈思宇,马愈卓,张雨佳,李尚泽,白彬沂,郝晓辉,罗本. 受治理原型知识迁移的无样本终身行人再识别[J/OL]. 中国图象图形学报. DOI: 10.11834/jig.260402.)

0 引 言

行人再识别(person re-identification, person ReID)旨在跨不相交的摄像机检索同一行人,是智能视频分析的关键技术,在安防布控、视频监控、智慧城市以及失踪人口查找等场景中具有重要的应用价值。近年来,行人再识别的研究已从传统的闭集身份匹配,逐步拓展到跨模态、低分辨率、朝向感知,以及预训练大模型驱动等多个方向(Zhang等,2023; Jiang和Ye, 2023; Yang等, 2023; Yang等, 2025; Feng等, 2025)。然而,真实部署环境在成像尺度、模态与场景条件上往往存在显著差异,导致明显的跨域分布偏移,这类差异难以通过单一静态模型一次性覆盖(Hu等, 2024; Sun等, 2023; Yan等, 2024; Li等, 2025)。

随着系统长期运行,摄像机、光照、背景乃至待识别身份都会持续更替,这要求模型能够不断学习新场景,同时不遗忘已有知识。针对这一要求,终身行人再识别(lifelong person re-identification, LReID)应运而生,其目标是使得行人再识别模型在顺序到来的领域上持续学习,并保持此前已获得的判别性知识(Pu等, 2021; Wu和Gong, 2021; Zhou等, 2025)。终身行人再识别面临的核心挑战是灾难性遗忘,即模型在学习新阶段的行人数据时,会削弱其早期阶段已学习到的身份区分能力(Li和Hoiem, 2018)。由于行人具有类内变化大、类间差异细微的特点,灾难性遗忘现象在行人再识别任务中尤为严重。基于回放的终身行人再识别方法通过存储历史图像来缓解遗忘(Wu和Gong, 2021),但行人图像的长期存储往往与隐私保护、存储成本及合规要求相冲突。正因如此,在每个阶段结束后均无法访问历史原始图像的无样本(exemplar-free)终身行人再识别,已成为一个重要且更具现实意义的研究方向(Pu等, 2021; Xu等, 2024a, 2024b)。

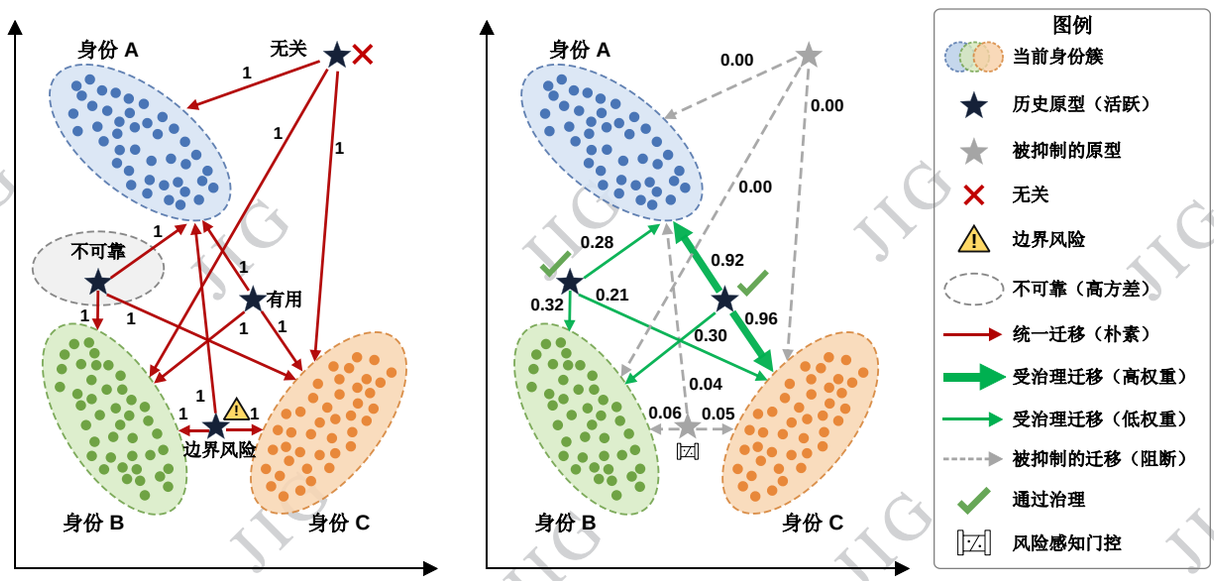
近期无样本终身行人再识别方法通常通过保存和迁移历史知识,而不是回放历史图像,来缓解灾难性遗忘。其中,基于原型记忆的方法将历史身份信息压缩为身份原型。训练新阶段时,历史身份原型作为替代性的结构知识,用于保持旧身份的判别边界或指导当前特征学习(Asadi等, 2023; Shi和Ye, 2023; Li等, 2024)。此外,基于知识蒸馏的方法通过对分类预测得分、特征、图像块或关系矩阵施加软约束,使当前模型与以往模型保持一致,从而保存旧知识(Hinton等, 2015; Pu等, 2021; Xu等, 2024a, 2024b; Cui等, 2024)。如图1(a)所示,这些方法通常将历史原型作为静态且同等可信的知识源重复使用:以分布感知知识对齐与原型化(distribution-aware knowledge aligning and prototyping, DKP++)方法(Zhou等, 2025)为例,其采用统一的等权原型迁移策略,图中各迁移路径的权重均为1。然而,随着学习阶段的不断累积,历史原型可能在领域相关性、样本支持与边界风险上与当前学习阶段的行人数据存在显著差异,特征空间漂移也可能使旧模型提取

© 中国图象图形学报版权所有

的原型与当前表示空间失配(Li 等,2024;Zhou 等,2025)。这些差异和不匹配会导致原型与当前样本的相似度估计以及原型诱导的关系构造变得不可靠,使得潜在有用的历史原型也可能被错误地分配到当前阶段特征区域,从而提供误导性的迁移信号。

进一步分析可以发现,统一复用历史原型主要带来三方面问题。其一是相关性问题:部分原型来自与当前小批量相距较远的领域,其根本原因在于不同领域的数据分布存在显著差异,这类无关原型几乎无法提供可迁移的身份结构,反而会向当前训练注入跨域噪声。其二是可靠性问题:部分原型由过少样本估计得到或携带异常方差,其本质是统计估计不充分,这会使原型自身带有噪声,从而产生不稳定甚至错误的监督信号。其三是边界风险问题:部分原型同时靠近多个当前身份中心、处于决策边界附近,其根源在于不同身份在特征空间中高度混淆,若强制迁移会进一步模糊身份边界。若对所有原型一视同仁地迁移,上述无关、含噪和边界风险知识将被一并注入,最终既削弱对旧领域判别知识的保持,又干扰对新领域可靠身份表示的学习。

由此可见,无样本终身行人再识别不仅需要构



(a)基线方法 (b)GPKT
((a) Baseline; (b) GPKT)

图 1　GPKT 与传统原型迁移方法的对比

Fig. 1　Comparison between GPKT and conventional prototype transfer methods

建原型,更需要对原型进行治理:只有当历史知识与当前领域相关、在统计上可靠、且相对当前身份边界安全时,才应被迁移。因此,本文提出受治理原型知识迁移(governed prototype knowledge transfer, GPKT)框架。如图 1(b)所示,GPKT 依据领域相关性、统计可靠性和边界安全性对候选历史原型进行差异化治理,只允许通过治理的原型影响当前身份关系,并用治理分数调节其贡献:可信原型获得较高迁移权重,不可靠原型被降权,边界风险原型受到强抑制,无关原型的迁移路径则被阻断。因此,历史原型在当前训练中的使用从无差别知识迁移转变为选择性、加权且安全的知识迁移。在此基础上,本文将原型的相关性、可靠性与边界风险纳入统一的治理框架,使历史知识的迁移更加可控、可靠,并在不增加推理开销的前提下兼顾旧领域保持与新领域适应。本文主要贡献概括如下:

1)针对历史原型被无差别复用的问题,本文将原型治理引入无样本终身行人再识别,以选择性、可靠且边界安全的迁移替代统一复用,从机制上避免无关、含噪和边界风险知识污染当前训练。

2)针对隐私与存储约束下历史知识难以保存的问题,本文提出层次化分布原型记忆,在不保留任何历史图像的前提下,以身份级分布原型、领域级描述符和样本支持统计量刻画历史知识,为后续的原型

© 中国图象图形学报版权所有

选择与治理提供依据。

3)针对历史原型与当前特征空间失配的问题，本文提出相关性引导的原型选择与校准，仅检索与当前小批量相关的历史领域和身份原型，并通过残差矩匹配缓解特征空间漂移，提升所迁移知识的相关性与可对齐性。

4)针对原型可靠性参差与边界混淆的问题，本文提出可靠性与风险感知原型治理，依据样本支持、方差稳定性、领域相关性和边界风险为原型分配治理权重，抑制不可靠或边界风险迁移。

5)针对历史身份关系结构难以在无样本条件下迁移的问题，本文提出受治理关系蒸馏，将治理权重注入样本到原型的分配，并以原型诱导的关系分布作为教师迁移历史身份关系，无需跨领域的显式身份对应。

6)在两种标准训练顺序下，GPKT在50层残差网络(50-layer residual network，ResNet-50)骨干设置中取得最佳的已见域平均与未见域平均的平均精度均值(mean average precision，mAP)/Rank-1准确率，且不引入额外的推理阶段原型开销；此外，本文通过逐阶段评估以及平均遗忘率、后向迁移和最终保持率等指标，直接分析GPKT在旧领域保持与新领域适应两方面的表现。

1 方 法

1.1 整体概述

本文聚焦于无样本终身行人再识别，其中模型在一系列有标注行人再识别领域$\mathbf{D}=\{\mathbf{D}_1,\mathbf{D}_2,\cdots,\mathbf{D}_t\}$上顺序训练。在训练阶段$t$，只有当前域数据集$\mathbf{D}_t$可用，而此前阶段的原始图像不能被存储或回放。令$f_{\theta_t}$表示参数为$\theta_t$的当前行人再识别模型，令$\mathbf{c}_i=f_{\theta_t}(\mathbf{x}_i)\in\mathrm{R}^d$表示图像$\mathbf{x}_i$的特征表示，其中$d$为特征维度。目标是在学习当前领域判别性表示的同时，保持此前领域中学到的身份判别能力。

GPKT的动机在于，历史原型不应被当作静态且同等可信的参考而无差别复用，而应区分对待。在终身行人再识别中，不同历史原型在领域相关性、统计可靠性与边界风险上存在显著差异，且特征空间漂移会进一步加剧其与当前表示空间的失配。以同等重要性迁移所有原型，会注入无关、含噪和边界风险的监督信号，混淆当前身份边界，并同时损害行人再识别模型对旧领域保持与新领域适应的能力。

如图2所示，GPKT在标准ReID特征提取主干之上构建受治理的原型迁移机制，整体遵循“选择—校准—治理—蒸馏”的流程。首先，层次化分布原型记忆在每个阶段结束后存储层次化的身份级分布原型、领域级描述符和样本支持统计量，从而在不保留原始行人图像的前提下，在身份和领域两个层面保存历史知识；其次，相关性引导的原型选择与校准依据特征中心和方差一致性检索与当前小批量最相关的前K个(top-K)历史领域(K为历史领域选择数)，保留最接近当前身份中心的前M个(top-M)身份原型(M为身份原型选择数)，并通过残差矩匹配校准所选原型以减小特征空间漂移；随后，可靠性与风险感知原型治理从样本支持、方差稳定性、领域相关性和边界风险四个方面评估每个校准后的原型，赋予治理权重以抑制不可靠或边界风险原型、突出可信原型；最后，受治理关系蒸馏将治理权重作为先验注入样本到原型的分配，构造原型诱导的教师关系分布，并通过KL散度(Kullback-Leibler divergence，KL)将当前特征关系分布与之对齐，从而以选择性、加权的方式迁移历史身份关系结构。上述所有原型相关操作均仅在训练阶段执行，推理阶段GPKT仍为标准的ReID特征提取器，不引入额外开销。下面各小节将依次介绍这四个模块。

1.2 层次化分布原型记忆

在无样本设定下，阶段结束后历史图像不可再访问，历史知识只能以紧凑统计形式保存；而仅保留单一身份中心会丢失类内分布、无法支撑后续的方差校准与可靠性评估。为此，第h个训练阶段($h<t$)结束后GPKT不存储来自$\mathbf{D}_h$的任何图像，而是将历史知识压缩为一个层次化的记忆单元$\mathbf{M}_h$。它以身份级分布原型保留类内变化、以领域级描述符刻画领域分布差异、以样本支持计数记录估计充分程度，从而在满足隐私与存储约束的同时，分别为后续的原型校准、相关性选择与可靠性治理提供直接依据：

$$\mathbf{M}_h=\{\mathbf{P}_h,\boldsymbol{\delta}_h,\mathbf{n}_h\},\quad h<t \tag{1}$$

式中，$\mathbf{M}_h$为第h个历史阶段的记忆单元；$\mathbf{P}_h$为身份级分布原型集合；$\boldsymbol{\delta}_h$为领域级分布描述符；$\mathbf{n}_h=\{n_j^h\}$为样本支持计数数组，其中n_j^h为历史领域h中身份j的支持样本数；h和t分别表示历史阶段索引和当前训

© 中国图象图形学报版权所有

练阶段索引。

对于每个身份$j \in \mathbf{Y}_h$，其中$\mathbf{Y}_h$表示历史领域$\mathbf{D}_h$的身份集合，本文存储一个身份级分布原型：

$$\mathbf{p}_j^h = \left(\boldsymbol{\mu}_j^h, \left(\boldsymbol{\sigma}_j^h\right)^2\right) \tag{2}$$

式中，$\mathbf{p}_j^h$为历史领域$\mathbf{D}_h$中身份j的分布原型；$\boldsymbol{\mu}_j^h$为该身份的特征均值向量；$\boldsymbol{\sigma}_j^h$为类内特征标准差向量，$\left(\boldsymbol{\sigma}_j^h\right)^2$表示类内特征逐维方差向量；两个统计量均由阶段固定模型f_{θ_h}计算得到。

此外，每个历史领域都有一个领域级描述符：

$$\boldsymbol{\delta}_h = \left(\bar{\boldsymbol{\mu}}_h, \bar{\boldsymbol{\sigma}}_h^2\right) \tag{3}$$

式中，$\boldsymbol{\delta}_h$为历史领域$\mathbf{D}_h$的分布描述符；$\bar{\boldsymbol{\mu}}_h$表示该领域所有图像的全局特征均值向量，$\bar{\boldsymbol{\sigma}}_h^2$表示该领域所有图像的全局特征方差向量。

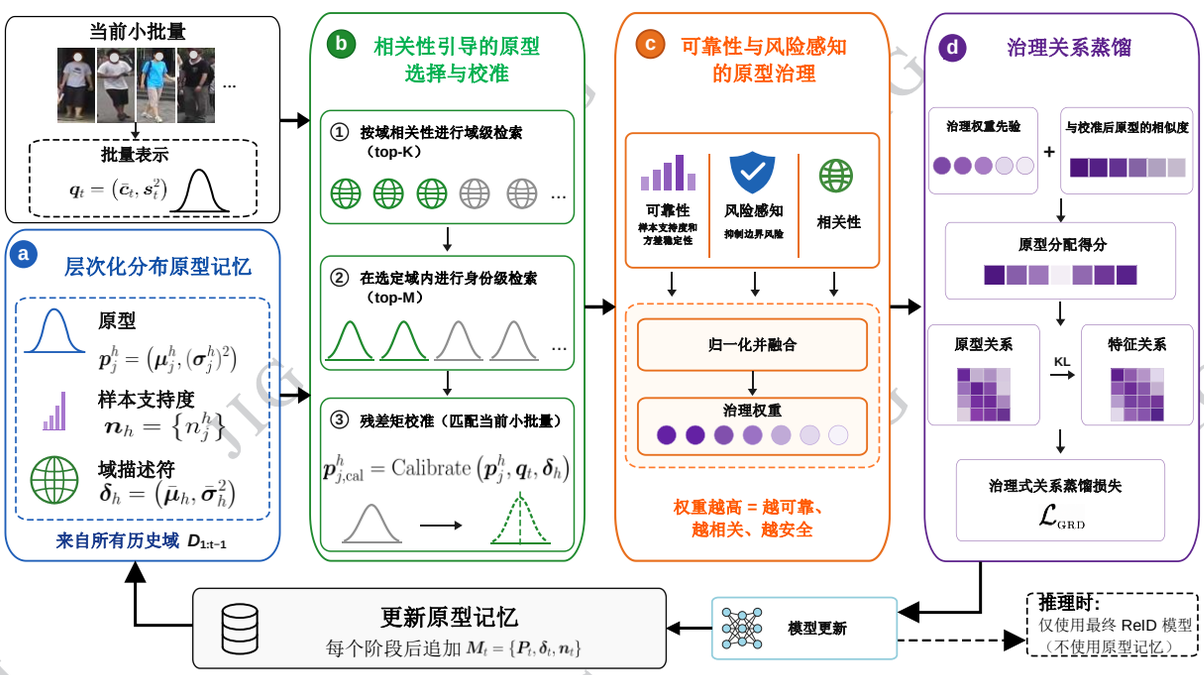
1.3 相关性引导的原型选择与校准

由于历史原型来自分布差异较大的多个领域，不加选择地全部使用既会随阶段累积而增大开销，又会因无关原型引入跨域噪声。为此，GPKT采用“先领域、后身份”的由粗到细选择：先检索与当前小批量相关的历史领域，再筛选其中靠近当前特征区域的身份原型，从而缩小搜索空间、在源头抑制跨域噪声。考虑到原型由早期固定模型提取、与当前空间未必严格对齐，而模型跨阶段的平滑更新使嵌入空间保持粗粒度几何连续性，GPKT仅用相似度做粗略检索，并将精确对齐与可靠性判定留给随后的校准与治理，以兼顾选择效率与迁移可靠性。

给定当前小批量特征$\mathbf{F}_c = \{\mathbf{c}_i\}_{i=1}^B$（$B$为小批量大小），本文用特征均值和方差概括当前小批量分布：

$$\mathbf{q}_t = \left(\bar{\mathbf{c}}_t, \mathbf{s}_t^2\right) \tag{4}$$

式中，$\mathbf{q}_t$为当前小批量的分布概要，$\bar{\mathbf{c}}_t = \frac{1}{B}\sum_{i=1}^B \mathbf{c}_i$，$\mathbf{s}_t^2$为小批量内逐维特征方差。均值刻画当前小批量的粗略特征位置，方差描述其逐维特征离散度，为域相关性估计与原型校准提供紧凑的分布概要。



((a) hierarchical distribution prototype memory; (b) relevance-guided prototype selection and calibration; (c) reliability- and risk-aware prototype governance; (d) governed relation distillation)

图2 GPKT总体框架

Fig. 2 Overall framework of GPKT

1.3.1 领域级选择

如果某个历史领域的特征中心接近当前小批量，并且其特征方差也与当前分布一致，则认为该历史领域相关。本文定义领域相关性分数为：

$$S_h^{\mathrm{dom}} = \mathrm{Norm}_{\mathbf{H}}\left(\cos\left(\bar{\mathbf{c}}_t, \bar{\boldsymbol{\mu}}_h\right)\right) - \mathrm{Norm}_{\mathbf{H}}\left(D_{\mathrm{var}}\left(\mathbf{q}_t, \boldsymbol{\delta}_h\right)\right) \tag{5}$$

式中，$\mathrm{Norm}_{\mathbf{H}}$表示在历史领域集合$\mathbf{H}$上的零均值、单位方差归一化，$\cos(\cdot,\cdot)$为余弦相似度函数，$D_{\mathrm{var}}(\cdot,\cdot)$

© 中国图象图形学报版权所有

为对数方差差异函数，计算如下：

$$D_{\mathrm{var}}(\mathbf{q}_t,\boldsymbol{\delta}_h)=\frac{1}{d}\left\|\log\left(\mathbf{s}_t^2+\quad\right)-\log\left(\bar{\boldsymbol{\sigma}}_h^2+\quad\right)\right\|_1 \tag{6}$$

式中，$\quad=10^{-6}$用于数值稳定性，$\|\cdot\|_1$表示L1范数。S_h^{dom}的第一项偏好特征中心相似的领域，第二项惩罚特征扩散不一致的领域。本文根据S_h^{dom}选择前K个历史领域，并将其记为$\mathbf{H}_K$。

1.3.2　身份级选择

检索相关领域后，本文进一步选择靠近当前特征区域的身份原型。令$\mathbf{Y}_{\mathrm{B}}$表示当前小批量中的身份集合，对于每个身份$y\in\mathbf{Y}_{\mathrm{B}}$，本文对属于$y$的样本特征求平均，得到其小批量身份中心$\mathbf{m}_y$。历史原型$(h,j)$与当前小批量之间的相关性为：

$$S_j^{\mathrm{id},h}=\max_{y\in\mathbf{Y}_{\mathrm{B}}}\cos\left(\boldsymbol{\mu}_j^h,\mathbf{m}_y\right) \tag{7}$$

该分数不假设历史身份与当前身份为同一人；在所选领域中保留前M个身份原型，集合记为$\mathbf{I}_{K,M}$。

1.3.3　原型校准

尽管所选原型相关，但它们由旧模型提取，可能与当前特征空间不对齐。因此，本文利用当前批统计量对其进行校准。对于历史领域h，本文计算基于方差的尺度比：

$$\mathbf{r}_h=\mathrm{clamp}_{[0.5,2.0]}\left(\sqrt{\frac{\mathbf{s}_t^2+\quad}{\bar{\boldsymbol{\sigma}}_h^2+\quad}}\right) \tag{8}$$

式中，clamp为截断函数，将尺度比限制在[0.5, 2.0]内；无校准情形$\mathbf{r}_h=1$，该约束用于防止极端缩放。

为避免符号混淆，本文用下标cal、mid分别表示校准和中点原型。随后，本文通过保留身份原型在历史领域中的相对位置，将每个历史原型对齐到当前小批量分布：

$$\boldsymbol{\mu}_{j,\mathrm{cal}}^h=\bar{\mathbf{c}}_t+\mathbf{r}_h\quad\left(\boldsymbol{\mu}_j^h-\bar{\boldsymbol{\mu}}_h\right) \tag{9}$$

$$\left(\boldsymbol{\sigma}_{j,\mathrm{cal}}^h\right)^2=\mathbf{r}_h^2\quad\left(\boldsymbol{\sigma}_j^h\right)^2 \tag{10}$$

式中，$\quad$表示逐元素乘法。项$\boldsymbol{\mu}_j^h-\bar{\boldsymbol{\mu}}_h$保留身份$j$在历史领域$h$内部的相对位置，而$\bar{\mathbf{c}}_t$与$\mathbf{r}_h$将该相对位置平移并重新缩放，以适应当前小批量。图2中记$\mathbf{p}_{j,\mathrm{cal}}^h=\left(\boldsymbol{\mu}_{j,\mathrm{cal}}^h,\left(\boldsymbol{\sigma}_{j,\mathrm{cal}}^h\right)^2\right)$为校准后的身份分布原型，Calibrate(·)表示式(8)—(10)所述的残差矩匹配校准过程。

由于小批量统计可能存在噪声，本文采用中点更新，以避免根据潜在不稳定的小批量估计对历史原型进行过度校正：

$$\boldsymbol{\mu}_{j,\mathrm{mid}}^h=\frac{\boldsymbol{\mu}_j^h+\boldsymbol{\mu}_{j,\mathrm{cal}}^h}{2} \tag{11}$$

$$\left(\boldsymbol{\sigma}_{j,\mathrm{mid}}^h\right)^2=\mathrm{clip}_\sigma\left(\frac{\left(\boldsymbol{\sigma}_j^h\right)^2+\left(\boldsymbol{\sigma}_{j,\mathrm{cal}}^h\right)^2}{2}\right) \tag{12}$$

式中，$\mathrm{clip}_\sigma(\cdot)$对每个方差维度施加下界裁剪，以避免退化方差，即$\mathrm{clip}_\sigma(v)=\max(v,\quad)$，$v$为任一维度上的方差值。校准后的原型$\left(\boldsymbol{\mu}_{j,\mathrm{mid}}^h,\left(\boldsymbol{\sigma}_{j,\mathrm{mid}}^h\right)^2\right)$随后用于原型治理和关系蒸馏。

1.4　可靠性与风险感知原型治理

原型选择保证了相关性、校准缓解了空间失配，但都未回答被选中的原型是否值得信赖：由少量样本估计的原型含噪，靠近多个身份中心的原型会模糊边界，因此仅靠选择与校准不足以保证迁移质量。为此，GPKT进一步以原型治理判断每个原型应在多大程度上影响当前训练，并用连续权重而非硬性取舍来调节其贡献，从而抑制含噪与边界风险原型、突出可信原型。治理权重由反映历史领域相关性的领域级先验，以及反映原型自身是否可靠且边界安全的原型级分数两部分构成，兼顾粗粒度效率与细粒度区分。

1.4.1　原型可靠性

可靠的原型应具有充足的样本支持与稳定的方差。由过少样本估计或方差异常的原型可能提供带有噪声的监督信号。本文首先对每个所选原型的样本数进行归一化：

$$\bar{n}_j^h=\frac{n_j^h}{n_{\max}+} \tag{13}$$

式中，n_j^h为身份原型j在历史领域h中的支持样本数，$n_{\max}$为所选领域中前M个身份原型组成的集合$\mathbf{I}_{K,M}$中，所有原型的最大支持样本数；归一化后的样本支持计数用于可靠性估计。

为度量方差稳定性，本文计算经校准原型的平均对数方差：

$$u_j^h=\frac{1}{d}\sum_{l=1}^{d}\log\left(\left(\sigma_{j,l,\mathrm{mid}}^h\right)^2+\quad\right) \tag{14}$$

式中，$\left(\sigma_{j,l,\mathrm{mid}}^h\right)^2$表示来自历史域$h$的第$j$个身份校准原型在第$l$个特征维度上的方差，$l$为特征维度索引，$d$为特征维度数。因此，$u_j^h$通过对所有特征维度上的对数方差求平均，刻画该校准历史原型整体的方差稳定性。

© 中国图象图形学报版权所有

然后，将原型可靠性分数定义为：

$$a_j^h = \sqrt{\bar{n}_j^h} \cdot \exp\left(-\left| \mathrm{Norm}_{\mathbf{I}_{K,M}}\left(u_j^h\right) \right| \right) \tag{15}$$

式中，$\bar{n}_j^h$表示归一化后的样本支持度，$\mathrm{Norm}_{\mathbf{I}_{K,M}}(\cdot)$表示在所选候选原型集合$\mathbf{I}_{K,M}$上进行的零均值、单位方差归一化，$\exp(\cdot)$为指数函数。该可靠性分数随样本支持度增加而增大，同时对偏离候选原型整体方差水平的异常原型进行惩罚。由于采用绝对标准化偏差，无论平均对数方差显著偏大或偏小，均会降低其方差稳定性权重；当原型的方差统计接近候选集合的典型水平时，该项取得较高值。因此，该设计衡量的是方差统计的稳定性与异常程度，而非简单地偏好方差越小越好。

1.4.2　边界安全迁移

靠近当前特征区域的历史原型更可能提供有用的关系信息。然而，如果某个历史原型同时接近多个当前身份中心，则它可能位于当前身份决策边界附近，从而在迁移过程中模糊不同身份之间的判别边界。因此，本文同时考虑历史原型与当前身份区域的接近性以及其潜在边界模糊性。

对于每个校准后的历史原型，本文首先计算其与当前小批量中所有身份中心的最大相似度，并将对应的最近身份记为y^*：

$$q_j^h = \max_{y \in \mathbf{Y}_\mathrm{B}} \cos\left(\boldsymbol{\mu}_{j,\mathrm{mid}}^h, \mathbf{m}_y\right) \tag{16}$$

$$y^* = \arg\max_{y \in \mathbf{Y}_\mathrm{B}} \cos\left(\boldsymbol{\mu}_{j,\mathrm{mid}}^h, \mathbf{m}_y\right) \tag{17}$$

如果当前小批量中包含多个身份，即$\left|\mathbf{Y}_\mathrm{B}\right| > 1$时，进一步计算该历史原型到最近身份中心与第二近身份中心之间的相似度间隔：

$$b_j^h = q_j^h - \max_{y \neq y^*} \cos\left(\boldsymbol{\mu}_{j,\mathrm{mid}}^h, \mathbf{m}_y\right) \tag{18}$$

式中，较大的q_j^h表明该历史原型位于当前特征区域附近，可能提供有效的关系知识；较大的b_j^h表明该原型主要靠近一个当前身份中心，而不是同时靠近多个身份中心。相反，较小的b_j^h说明该原型可能处在多个当前身份之间的边界区域，容易引入边界模糊或错误迁移。

为了对上述两个因素进行平滑建模，本文定义接近性门控和间隔门控：

$$\eta_j^h = \mathrm{sigmoid}\left(\frac{q_j^h}{\tau_\mathrm{r}}\right) \tag{19}$$

$$\kappa_j^h = \mathrm{sigmoid}\left(\frac{b_j^h}{\tau_\mathrm{r}}\right) \tag{20}$$

式中，η_j^h表示接近性门控，用于衡量历史原型是否靠近当前特征区域；κ_j^h表示间隔门控，用于衡量该原型是否远离当前身份边界；τ_r为风险门控温度；$\mathrm{sigmoid}(\cdot)$为标准sigmoid函数。

如果当前小批量中仅包含一个身份，则不存在第二近身份中心，因而无法估计身份间边界模糊性。在这种情况下，本文将间隔门控设为中性值$\kappa_j^h = 1$。这一设置并不表示该原型一定是边界安全的，而是表示在无法估计边界间隔时，不通过间隔门控对其进行额外惩罚，使边界安全性主要由接近性门控控制。最终，边界安全得分定义为：

$$s_{j,\mathrm{safe}}^h = \eta_j^h \kappa_j^h \tag{21}$$

因此，只有当历史原型既靠近当前特征区域，又不会同时接近多个当前身份中心时，才会获得较高的边界安全得分。该得分随后与原型可靠性得分共同用于计算治理权重，从而抑制边界风险较高的历史原型对当前关系蒸馏过程的影响。

1.4.3　治理权重

为同时刻画每个历史原型的统计可信性与边界安全迁移性，本文将可靠性分数和边界安全分数以乘法方式整合为原型级治理分数：

$$\psi_j^h = a_j^h \cdot \left(1 + s_{j,\mathrm{safe}}^h\right) \tag{22}$$

式中，“·”表示乘法关系。由式(15)及式(19)—(21)可知，可靠性分数a_j^h满足$a_j^h \in (0,1)$，边界安全分数$s_{j,\mathrm{safe}}^h$满足$0 < s_{j,\mathrm{safe}}^h < 1$，因此乘性因子$(1 + s_{j,\mathrm{safe}}^h) \in (1,2)$，治理分数$\psi_j^h$满足$\psi_j^h \in (0,2)$。其中，$a_j^h$是用于表征原型相对可靠程度的连续权重，其数值越大，表示该原型具有更充分的样本支持且方差统计更接近候选原型集合的典型水平。随后，本文在每个所选域内对原型级治理分数进行归一化：

$$\pi_j^h = \frac{\psi_j^h}{\sum_{j' \in \mathbf{I}_{K,M}^h} \psi_{j'}^h +} \tag{23}$$

式中，$\mathbf{I}_{K,M}^h$表示来自历史域h的所选候选原型集合，j'为该集合中的虚拟求和索引，用于遍历同一历史域内的所有被选中原型。式(23)在每个历史域内部对治理分数单独归一化，使域内归一化权重π_j^h满足$0 < \pi_j^h < 1$，且同一域内所有π_j^h之和近似为1(　仅用于数值稳定)。因此，π_j^h表示原型在所属域内的相对治理重要性，域内归一化只决定该域内部各原

© 中国图象图形学报版权所有

型之间的贡献比例，避免候选原型数量较多的历史域仅因原型数更多而获得更大的总体权重；不同历史域之间的总体贡献由后续域级先验α_h控制。

进一步地，利用域级选择步骤中定义的域相关性分数S_h^{dom}，本文在所选历史域集合$\mathbf{H}_K$上计算域级先验：

$$\alpha_h = \frac{\exp\left(S_h^{\mathrm{dom}}\right)}{\sum_{h' \in \mathbf{H}_K} \exp\left(S_{h'}^{\mathrm{dom}}\right)} \tag{24}$$

式中，h'为所选历史域集合$\mathbf{H}_K$中的虚拟求和索引，用于遍历所有被选历史域。式(24)通过softmax对所选历史域的相关性分数进行归一化，得到域级先验α_h，其取值满足$0 < \alpha_h < 1$，且所有被选历史域的α_h之和为1。α_h表示历史域h相对于其他被选历史域的整体相关性，并作为域级缩放因子控制该域内所有原型的总体贡献：历史域与当前小批量越相关，α_h越大，其域内原型在最终治理权重中整体获得更高权重；反之则被整体抑制。

最终，第h个历史域中第j个原型的治理权重定义为：

$$\omega_j^h = \alpha_h \pi_j^h \tag{25}$$

式中，$\omega_j^h = \alpha_h \pi_j^h$以“域级先验×域内原型权重”的层级形式得到最终治理权重。α_h决定历史域h的总体贡献，π_j^h决定该贡献在域内各候选原型之间的分配。因此，同一历史域内各ω_j^h之和近似等于α_h，而全部被选原型的治理权重之和近似为1。由此，一个历史原型只有在所属历史域与当前数据分布较为相关、原型本身具有较高统计可靠性且边界迁移风险较低时，才会获得较大的治理权重。该设计能够实现对历史原型的软性抑制与加权利用，从而减少无关、不可靠或边界风险较高的历史知识对当前训练的负面影响。

1.5 受治理关系蒸馏

在无样本设定下，历史身份与当前身份并非同一批人，不存在显式对应，直接做身份级对齐既无依据又会引入错误监督。为此，GPKT不强制跨领域身份匹配，而是以受治理原型为媒介迁移身份间的关系结构，即各样本在共享参考原型上的相对分布。其优势在于，关系结构无需身份对应即可刻画旧身份在特征空间中的相对组织，从而在无标签对应时迁移历史判别结构；且参考原型均经治理，治理权重沿关系构造传入蒸馏，自动抑制无关、含噪与边界风险原型的影响。

具体来说，经治理后每个所选原型由其校准后的均值$\mathbf{g}_j^h = \boldsymbol{\mu}_{j,\mathrm{mid}}^h$表示，并用于引导当前小批量内的关系学习。本文将这些受治理原型展平并以$k = 1, \cdots, L$重新编号（其中$L = \left|\mathbf{I}_{K,M}\right|$），用$\mathbf{g}_k$与$\omega_k$分别表示第$k$个原型的代表向量与治理权重。对每个当前样本$\mathbf{c}_i$，本文计算其到原型$k$的分类预测得分：

$$_{ik} = \frac{\mathrm{sim}\left(\mathbf{c}_i, \mathbf{g}_k\right)}{\tau} + \log\left(\omega_k + \quad\right) \tag{26}$$

式中，$\mathrm{sim}(\cdot,\cdot)$是余弦相似度，$\tau$是关系蒸馏温度。第一项衡量特征相似度，第二项将治理权重作为先验注入。因此，可靠且安全的原型对关系构造贡献更强。

为得到当前样本对不同受治理历史原型的相对响应强度，本文对分配分类预测得分执行softmax操作，将其转换为归一化的样本—原型分配概率：

$$C_{ik} = \frac{\exp\left(\quad_{ik}\right)}{\sum_{k'=1}^{L} \exp\left(\quad_{ik'}\right)} \tag{27}$$

式中，k'是从1到L的虚拟原型索引。收集所有分配概率得到分配矩阵$\mathbf{C} \in \mathrm{R}^{B \times L}$，其第$(i,k)$个元素为$C_{ik}$，描述当前样本$i$与受治理历史原型$k$之间的关系。

基于该分配矩阵，本文构造当前样本之间的原型诱导关系分布：

$$\mathbf{R}_{\mathrm{proto}} = \rho\left(\frac{\mathbf{C}\mathbf{C}}{\tau} + \mathbf{A}\right) \tag{28}$$

式中，$\rho(\cdot)$表示按行softmax。这里使用相同的τ，是因为原型分配与关系构建是构造同一关系教师的两个相继步骤，使用同一温度可保持平滑尺度一致，而不引入额外的温度超参数。$\mathbf{A} \in \mathrm{R}^{B \times B}$是自关系掩码，其对角元素被设为很大的负值，非对角元素被设为零，从而在按行softmax时使每个样本不关注自身。因此，$\mathbf{R}_{\mathrm{proto}}$描述当前样本如何依据其对历史原型的共同响应而相互关联。

与此同时，本文基于当前小批量特征构建学生关系分布，用以表征当前模型在特征空间中学习到的样本间关系结构：

$$\mathbf{R}_{\mathrm{feat}} = \rho\left(\frac{\bar{\mathbf{F}}_{\mathrm{c}}\bar{\mathbf{F}}_{\mathrm{c}}}{\tau} + \mathbf{A}\right) \tag{29}$$

式中，$\bar{\mathbf{F}}_{\mathrm{c}} \in \mathrm{R}^{B \times d}$是按行归一化的当前特征矩阵。

最后，本文将原型诱导关系分布作为教师，将当

© 中国图象图形学报版权所有

前特征关系分布作为学生，计算受治理关系蒸馏（governed relation distillation，GRD）：

$$_{\mathrm{GRD}} = \mathrm{KL}\left(\mathrm{sg}\left(\mathbf{R}_{\mathrm{proto}}\right) \quad \mathbf{R}_{\mathrm{feat}}\right) \tag{30}$$

式中，KL(· ·)为KL散度，sg(·)表示停止梯度操作。由于历史身份与当前身份通常互不相交，该损失并不强制跨域身份匹配，而是通过正则化当前样本之间的关系来迁移历史身份结构知识。

1.6 训练目标与推理

GPKT的训练目标需同时满足两个要求：既要在当前领域学到具判别力的表示，又要保持此前阶段已获得的身份判别能力。为此，本文将标准行人再识别损失与受治理关系蒸馏损失相结合，其中身份分类损失 $_{\mathrm{ce}}$ 与三元组损失 $_{\mathrm{tri}}$ 负责当前领域的判别性学习（可塑性），受治理关系蒸馏损失 $_{\mathrm{GRD}}$ 则以历史身份关系结构约束当前模型、缓解灾难性遗忘（稳定性）。二者相加，使模型在可塑性与稳定性保持之间取得平衡。总训练损失 $_t$ 定义如下：

$$_t = \quad _{\mathrm{ce}} + \quad _{\mathrm{tri}} + \mathrm{I}[t > 1]\lambda \quad _{\mathrm{GRD}} \tag{31}$$

式中，指示函数$\mathrm{I}[t > 1]$表明蒸馏项仅在历史记忆可用时（即自第二阶段起）才被激活：首个阶段尚无历史知识，此时仅以标准ReID损失训练，从而避免引入无意义的正则。权重λ统一调节历史关系迁移的强度、平衡保持与适应，本文设$\lambda = 0.7$。

每个阶段结束后，本文以该阶段固定的模型构造新的记忆单元$\mathbf{M}_t$并追加到记忆池，供后续阶段的原型选择、校准与治理使用，从而实现历史知识的持续累积。值得强调的是，上述所有原型相关操作与蒸馏损失均仅作用于训练阶段；推理时GPKT为标准的ReID特征提取器，不引入任何额外的存储或计算开销。

2 实 验

2.1 实验设置

2.1.1 数据集与训练顺序

本文在标准终身行人再识别基准上评估GPKT（Pu等，2021；Zhou等，2025），该基准包含12个行人再识别数据集。Market-1501、CUHK-SYSU、DukeMTMC、MSMT17和CUHK03被用作终身训练与评估的已见领域。CUHK01、CUHK02、VIPeR、PRID、iLIDS、GRID和SenseReID被用作泛化评估的未见领域。本文遵循两种标准训练顺序：

顺序-1：依次为Market-1501、CUHK-SYSU、DukeMTMC、MSMT17、CUHK03。

顺序-2：依次为DukeMTMC、MSMT17、Market-1501、CUHK-SYSU、CUHK03。

2.1.2 评价指标

本文采用行人再识别领域常用的平均精度均值（mean average precision，mAP）和Rank-1准确率（R1）作为检索性能评价指标。已见域平均表示所有已见领域的平均性能，未见域平均表示7个未见领域的平均性能。

2.1.3 对比方法

本文选取9种近期代表性无样本终身行人再识别方法作为对比方法，包括原型—样本关系蒸馏（prototype-sample relation distillation，PRD）（Asadi等，2023）、原型回忆与增强非对称知识聚合（prototype reminiscence and augmented asymmetric knowledge aggregation，PRAKA）（Shi和Ye，2023）、特征校准与分离（feature calibration and separation，FCS）（Li等，2024）、一致区域特征学习（consistent region feature learning，ConRFL）（Huang等，2023）、记忆与泛化框架（memorizing and generalizing framework，MEGE）（Pu等，2023）、持续知识净化（continual knowledge purification，CKP）（Xu等，2024a）、长短期知识巩固（long short-term knowledge consolidation，LSTKC）（Xu等，2024b）、持续兼容表示（continual compatible representation，C^2R）（Cui等，2024）和分布感知知识对齐与原型化（distribution-aware knowledge aligning and prototyping，DKP++）（Zhou等，2025）。由于任务设定或监督信息并不一致，部分相关工作未直接纳入对比。具体而言，基于文本提示的终身行人再识别（lifelong person re-identification using text prompts，TP-LReID）方法（Liu等，2026）和提示锚定视觉—文本蒸馏（prompt-anchored vision-text distillation）方法（Wen等，2026）均引入文本提示或视觉—语言预训练知识，其额外模态监督和预训练先验明显不同于本文的纯视觉设定，因此不具备严格的同骨干、同监督公平比较条件。

2.1.4 实现细节

考虑到现有大部分终身行人再识别方法采用ResNet-50（He等，2016）作为骨干网络，本文首先使用ResNet-50作为GPKT的默认骨干网络，以便与现

© 中国图象图形学报版权所有

有方法进行同骨干网络设置下的公平比较。为进一步考察所提框架对不同视觉骨干网络的适用性，本文还将GPKT的骨干网络由ResNet-50替换为视觉Transformer（vision Transformer-base/16，ViT-B/16）（Dosovitskiy等，2021），记为GPKT-ViT；除骨干网络外，其他训练流程与方法设置保持一致。模型在第一个领域上训练80个epoch，并在每个后续领域上训练60个epoch。每个小批量包含从16个身份中采样的64张图像，每个身份4张图像。本文设置$K=3$、$M=12$、$\tau=0.10$、$\tau_r=0.20$、$\lambda=0.70$。采用随机梯度下降（stochastic gradient descent，SGD）优化，学习率为0.008，权重衰减为0.000 1。为保护隐私，行人图像均经过人脸遮挡，且该预处理统一作用于所有对比方法。所有对比方法均基于公开实现或依据原论文设置，并在本文统一实验协议下完成复现；所有结果均独立运行5次后取平均值。对于多场景特征分析，各方法在完全一致的样本上提取并L2归一化高维特征，采用一致的身份数量、点数、随机种子和可视化尺度以保证不同方法与不同场景之间的可比性；轮廓系数（silhouette coefficient，SC）、局部类间/类内分离比（separation ratio，Sep）和最大均值差异（maximum mean discrepancy，MMD）均在降维前的高维特征上计算，t分布随机邻域嵌入（t-distributed stochastic neighbor embedding，t-SNE）仅用于二维展示。

2.1.5　超参数选择与评估协议

为避免测试信息参与模型选择，本文采用严格的数据隔离与超参数锁定协议。所有超参数（K、M、τ、τ_r和λ）以及学习率、训练轮数等训练设置，均仅依据从已见训练域中划分出的独立验证集确定：在顺序-1下，从每个已见领域的训练集中按身份不重叠的原则随机留出10%的训练身份构成验证集；每组候选配置均完成完整五阶段验证流程后统一比较，比较依据仅为各阶段验证集上的已见域平均mAP整体表现，整个训练序列只确定一组默认配置，不在不同阶段分别选参或更换参数。默认配置一经确定即锁定，随后使用完整训练集重新训练；两种训练顺序以及ResNet-50与ViT-B/16两种骨干网络均沿用同一组默认超参数，不针对训练顺序或骨干网络单独调节。训练过程中不采用早停，每个阶段均训练固定轮数，并以该阶段最后一个epoch的模型进行评估和构造记忆单元。对比方法均采用其原论文或公开实现中推荐的超参数，同样不使用测试集进行调节。默认配置锁定之前，7个未见域测试集及各已见领域测试集均不参与模型选择、超参数调节或早停，也不作为候选配置的补充判据或并列判定依据；默认配置锁定之后，测试集仅用于结果报告、逐阶段评估以及第2.7节的事后稳健性分析，任何测试结果均不反向用于修改默认配置、训练设置或早停策略。

2.2　与近期方法的比较

表1和表2展示了GPKT与近期代表性终身行人再识别方法在两种训练顺序下的比较结果，其中最佳值和次佳值仅在采用ResNet-50骨干的方法中判定，分别以粗体和下划线表示。总体而言，在与现有方法相同的ResNet-50骨干网络下，GPKT在两种顺序中均取得最佳已见域平均和未见域平均mAP/R1性能。与平均性能最强的对比方法DKP++相比，GPKT在顺序-1上使已见域平均的mAP/R1分别提升3.1/2.7个百分点，并使未见域平均提升3.2/2.6个百分点。在顺序-2上，已见域平均提升3.0/2.9个百分点，未见域平均提升3.3/2.6个百分点。这些一致的平均性能提升表明，受治理原型迁移在最终阶段的已见域整体性能和跨领域泛化上均具有优势。需要说明的是，最终阶段的已见域平均性能尚不能区分旧领域保持与新领域适应各自的贡献，本文在第2.3节中通过逐阶段评估与遗忘指标对此进行直接分析。

在此基础上，GPKT-ViT进一步验证了所提框架在Transformer骨干网络上的适用性。结果显示，GPKT-ViT在顺序-1上将已见域平均的mAP/R1提升至60.1%/72.7%，未见域平均提升至70.7%/62.6%；在顺序-2上对应结果为60.0%/73.2%和70.1%/61.9%。与ResNet-50骨干网络下的GPKT相比，GPKT-ViT在两种顺序上已见域平均的mAP分别进一步提升1.5和1.5个百分点，未见域平均的mAP分别提升1.8和1.9个百分点，说明受治理原型迁移不仅适用于卷积骨干网络，也适用于Transformer骨干网络。

进一步观察各种对比方法在ResNet-50骨干网络下，在各个已见领域上的数据可以发现，GPKT相比DKP++的提升并非由单一数据集主导。在两种训练顺序下，GPKT在所有已见领域上均稳定优于DKP++。例如，在顺序-1上，GPKT在DukeMTMC和

© 中国图象图形学报版权所有

表1　顺序-1下与近期方法的性能比较

Table 1　Performance comparison with recent methods under Order-1

方法	发表	Market-1501		CUHK-SYSU		DukeMTMC		MSMT17		CUHK03		已见域平均		未见域平均	
		mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
PRD	ICML23	7.3	18.0	33.5	35.6	3.7	7.6	0.8	2.4	33.8	33.8	15.8	19.5	23.0	17.7
PRAKA	ICCV23	37.4	61.3	69.3	71.8	35.4	55.0	10.7	27.2	**54.0**	**55.6**	41.4	54.2	47.7	41.6
FCS	CVPR24	58.3	78.5	75.1	76.2	42.6	59.8	10.2	24.3	35.3	34.9	44.3	54.7	52.1	44.2
ConRFL	PR23	59.2	78.3	82.1	84.3	45.6	61.8	12.6	30.4	51.7	53.8	50.2	61.7	57.4	52.3
MEGE	TPAMI23	39.0	61.6	73.3	76.6	16.9	30.3	4.6	13.4	36.4	37.1	34.0	43.8	47.7	44.0
CKP	MM24	53.8	76.0	81.2	83.0	49.7	67.0	18.4	40.8	44.1	45.8	49.4	62.5	58.0	51.0
LSTKC	AAAI24	54.7	76.0	81.1	83.4	49.4	66.2	20.0	43.2	44.7	46.5	50.0	63.1	57.0	49.9
C^2R	CVPR24	**69.0**	**86.8**	76.7	79.5	33.2	48.6	6.6	17.4	35.6	36.2	44.2	53.7	48.7	41.9
DKP++	TPAMI25	63.9	83.6	83.7	85.3	56.7	74.1	26.5	52.6	46.5	47.6	55.5	68.6	65.7	58.5
GPKT	Ours	66.5	85.4	**86.2**	**88.1**	**60.1**	**77.4**	**29.8**	**56.0**	50.2	49.6	**58.6**	**71.3**	**68.9**	**61.1**
GPKT-ViT	Ours	68.1	86.4	87.5	89.0	61.8	78.5	31.6	58.2	51.4	51.3	60.1	72.7	70.7	62.6

注：加粗表示最优值，下划线表示次优值；最优值与次优值仅在采用ResNet-50骨干的方法中判定。

MSMT17上分别带来3.4/3.3和3.3/3.4个百分点的mAP/R1提升；在顺序-2上，在MSMT17和Market-1501上的提升分别达到3.4/3.2和3.5/2.5个百分点。这些结果说明，所提出的方法在不同领域以及终身序列的不同位置均能提供稳定收益。未见领域结果进一步表明，GPKT并未过拟合于已见训练领域。在相同骨干网络设置下，其在两种顺序中均取得最佳未见域平均性能，表明GPKT在跨领域泛化方面具有一致优势。

逐域结果进一步表明，顺序-1下GPKT在7个未见域上均取得mAP和R1正增益；顺序-2下则有6/7个未见域同时取得正增益，SenseReID上出现轻微回落。不同数据集上的收益幅度存在明显差异，而两种顺序的未见域平均性能仍保持稳定提升，说明整体收益来自多域共同贡献而非单一测试域拉动；具体逐域结果与主导分布偏移类型分析见第2.4节。

需要指出的是，GPKT并非在所有已见领域上均领先，逐数据集比较中有四个“域-顺序”位置未取得最优（涉及三种对比方法）。下面结合各方法设计特点给出机制层面的可能解释。首先，C^2R在顺序-1的Market-1501（mAP/R1为69.0%/86.8%，本文为66.5%/85.4%）与顺序-2的DukeMTMC（mAP为59.7%，本文为58.8%）上占优，而这两个数据集恰好分别是两种顺序中的首个训练域。C^2R以持续兼容表示为目标，其兼容性约束要求更新后的模型特征与历史模型特征在同一度量空间内可比，因此从机制上更偏向跨阶段表示稳定性；GPKT迁移的是样本间的关系结构而非特征的绝对位置，允许特征空间随新领域持续演化，这一设计差异可能是GPKT在最早阶段域上略有落后的原因之一；从整体结果看，在两种顺序下C^2R的已见域平均与未见域平均均低于GPKT。其次，PRAKA在CUHK03上于两种顺序下均优于本文方法（顺序-1为54.0%/55.6%对50.2%/49.6%，顺序-2为54.9%/56.6%对52.0%/55.0%）；同时，在顺序-1的CUHK03上，ConRFL也达到51.7%/53.8%，高于本文的50.2%/49.6%。CUHK03在两种顺序中均为最后一个训练域，其规模较小、检测框噪声大、分辨率偏低。PRAKA的原型增广与非对称知识聚合可能更有利于当前域拟合；ConRFL强调一致区域特征学习，可能对检测框噪声和局部区域扰动较明显的CUHK03具有一定适应性。相比之下，GPKT在末阶段需同时承受来自最多历史领域的受治理关系蒸馏约束（权重λ在各阶段固定），这可能在一定程度上限制其对末端小规模领域的专门拟合。GPKT更侧重于在连续学习过程中获得较为均衡的稳定性、可塑

© 中国图象图形学报版权所有

表2　顺序-2下与近期方法的性能比较

Table 2　Performance comparison with recent methods under Order-2

方法	发表	DukeMTMC		MSMT17		Market-1501		CUHK-SYSU		CUHK03		已见域平均		未见域平均	
		mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1	mAP	R1
PRD	ICML23	3.6	8.2	0.6	1.8	8.9	22.3	34.6	36.1	35.4	35.3	16.6	20.7	20.7	15.0
PRAKA	ICCV23	31.2	48.7	6.6	19.1	47.8	69.8	70.4	73.0	**54.9**	**56.6**	42.2	53.4	48.4	41.1
FCS	CVPR24	53.6	70.0	9.5	23.5	48.7	69.9	76.2	78.2	37.1	38.4	45.0	56.0	52.7	45.1
ConRFL	PR23	34.4	51.3	7.6	20.1	61.6	80.4	82.8	85.1	49.0	50.1	47.1	57.4	57.9	53.4
MEGE	TPAMI23	21.6	35.5	3.0	9.3	25.0	49.8	69.9	73.1	34.7	35.1	30.8	40.6	44.3	41.1
CKP	MM24	49.4	67.0	14.5	33.8	56.0	77.6	83.2	84.9	45.3	47.1	49.7	62.1	57.2	50.0
LSTKC	AAAI24	49.9	67.6	14.6	34.0	55.1	76.7	82.3	83.8	46.3	48.1	49.6	62.0	57.6	49.6
C^2R	CVPR24	**59.7**	75.0	7.3	19.2	42.4	66.5	76.0	77.8	37.8	39.3	44.6	55.6	48.5	41.7
DKP++	TPAMI25	55.7	73.9	22.2	47.8	65.9	85.6	84.4	86.5	49.5	51.7	55.5	69.1	64.9	57.9
GPKT	Ours	58.8	**76.5**	**25.6**	**51.0**	**69.4**	**88.1**	**86.8**	**89.2**	52.0	55.0	**58.5**	**72.0**	**68.2**	**60.5**
GPKT-ViT	Ours	60.4	77.6	27.6	53.2	70.9	89.0	88.1	90.1	53.2	56.0	60.0	73.2	70.1	61.9

注：加粗表示最优值，下划线表示次优值；最优值与次优值仅在采用ResNet-50骨干的方法中判定。

性与跨域泛化，而不是追求某一单独领域上的最高性能。

2.3　阶段性性能演化与遗忘分析

第2.2节报告的最终阶段已见域平均性能只能反映完整训练结束后的整体效果，既不能说明模型在每个阶段遗忘了多少，也不能区分性能优势来自旧领域保持、新领域适应还是二者共同作用。为此，本文在两种训练顺序下，于每个训练阶段结束后，使用该阶段的模型在截至当前阶段的全部已见领域测试集上进行评估，得到阶段性性能矩阵。记$a_{t,i}$为第t个阶段结束后模型在第i个已见领域$\mathbf{D}_i$上的mAP（$1 \leqslant i \leqslant t \leqslant T$），$T=5$为阶段总数。基于该矩阵，本文采用以下指标。阶段平均性能A_t刻画第t个阶段结束时模型在全部已学领域上的整体性能，其中A_T即第2.2节中的已见域平均mAP：

$$A_t = \frac{1}{t}\sum_{i=1}^{t} a_{t,i} \tag{32}$$

学习精度（learning accuracy，LA）衡量各领域刚完成训练时的性能，用于反映新领域适应能力（可塑性）：

$$\mathrm{LA} = \frac{1}{T}\sum_{t=1}^{T} a_{t,t} \tag{33}$$

平均遗忘率（average forgetting，AF）衡量各旧领域从其历史最佳性能到最终性能的下降幅度（l为第i至第$T-1$阶段中的阶段索引），值越小表示遗忘越少：

$$\mathrm{AF} = \frac{1}{T-1}\sum_{i=1}^{T-1} \max_{l \in \{i,\cdots,T-1\}} \left(a_{l,i} - a_{T,i}\right) \tag{34}$$

后向迁移（backward transfer，BWT）衡量后续领域的学习对旧领域性能的影响，负值表示发生遗忘，值越大越好：

$$\mathrm{BWT} = \frac{1}{T-1}\sum_{i=1}^{T-1} \left(a_{T,i} - a_{i,i}\right) \tag{35}$$

最终保持率（retention ratio，RR）衡量各旧领域的最终性能相对其刚完成训练时性能的保持比例：

$$\mathrm{RR} = \frac{1}{T-1}\sum_{i=1}^{T-1} \frac{a_{T,i}}{a_{i,i}} \times 100\% \tag{36}$$

上述指标均基于mAP计算，其中LA反映新领域适应，AF、BWT和RR反映旧领域保持。DKP++按其公开实现在本文统一协议下复现，与GPKT采用相同的数据划分、训练轮数和评估方式，所有结果均为5次独立运行的平均值。两种训练顺序下的阶段性性能矩阵列于表3，遗忘与保持指标列于表4，阶段平均性能曲线如图3所示，各旧领域的遗忘量如图4所示。

由表3和表4可见，在顺序-1/顺序-2下，GPKT的平均遗忘率AF分别为5.85/5.90个百分点，较DKP++（7.78/7.90个百分点）分别降低1.93和2.00

© 中国图象图形学报版权所有

个百分点，即约2个百分点；BWT分别为−5.78/−5.90个百分点，最终保持率RR分别为90.53%/89.76%，均优于DKP++（BWT为−7.78/−7.90个百分点，RR为87.38%/85.74%）。需要说明的是，AF以各旧领域在第i至第$T-1$阶段中的最佳性能为参照，BWT则以各领域刚完成训练时的性能为参照；顺序-1下GPKT在CUHK-SYSU上的mAP于第3阶段达到最高（89.0%，高于第2阶段刚完成训练时的88.7%），因此该设置下AF与BWT的绝对值略有差异。与此同时，GPKT与DKP++的学习精度LA在顺序-1下分别为63.18%和61.68%，在顺序-2下分别为63.24%和61.86%，GPKT分别高出1.50和1.38个百分点，说明GPKT在减轻遗忘的同时也提升了新领域适应能力。

为进一步区分最终性能优势中旧领域保持与新领域适应各自的贡献，由式(32)、(33)和(35)可直接推得最终阶段已见域平均性能的恒等分解：

$$A_T = \mathrm{LA} + \frac{T-1}{T}\mathrm{BWT} \tag{37}$$

式中，LA项反映新领域适应，$(T-1)$BWT/T项反映旧领域保持。据此，顺序-1下GPKT相对DKP++的最终增益3.10个百分点可精确分解为适应项1.50个百分点与保持项0.8×2.00=1.60个百分点；顺序-2下的最终增益2.98个百分点（表2中四舍五入为3.0个百分点）可分解为适应项1.38个百分点与保持项1.60个百分点。由此可见，GPKT的最终性能优势来自旧领域保持与新领域适应的共同贡献，二者大体相当，其中旧领域保持略占多数（在两种顺序下分别约占52%和54%）。

图3给出了两种方法的阶段平均性能A_t随训练阶段的变化，图中点旁数字为GPKT与DKP++之差（百分点），GPKT以红色实线和圆形标记表示，DKP++以蓝色虚线和方形标记表示。两种方法在前两个阶段的A_t接近：顺序-1第1、2阶段GPKT分别低1.8和0.2个百分点，顺序-2前两个阶段二者相差不超过0.2个百分点。需要说明的是，第1阶段尚无历史记忆，GPKT的蒸馏项未被激活（见式(31)），仅以身份分类损失与三元组损失训练；而DKP++按其公开实现，在首阶段即包含其分布建模相关的训练目标。二者首阶段的目标函数与训练设置不同，因此第1阶段的差距反映的是初始模型的差异，而非原型治理的作用。从第3阶段起，GPKT的A_t持续高于DKP++，差距在最后一个阶段达到3.1和3.0个百分点。这说明GPKT的优势随历史领域的积累逐步显现。

表3　两种训练顺序下GPKT与DKP++的阶段性已见域mAP(%)

Table 3　Stage-wise seen-domain mAP (%) of GPKT and DKP++ under the two training orders

阶段	方法	D_1	D_2	D_3	D_4	D_5	A_t
顺序-1（D_1—D_5依次为Market-1501、CUHK-SYSU、DukeMTMC、MSMT17、CUHK03）							
1	DKP++	78.2	—	—	—	—	78.2
1	GPKT	76.4	—	—	—	—	76.4
2	DKP++	75.0	87.9	—	—	—	81.5
2	GPKT	73.9	88.7	—	—	—	81.3
3	DKP++	69.9	84.6	64.5	—	—	73.0
3	GPKT	72.6	89.0	66.6	—	—	76.1
4	DKP++	67.8	84.6	60.4	31.3	—	61.0
4	GPKT	68.7	87.2	62.5	34.0	—	63.1
5	DKP++	63.9	83.7	56.7	26.5	46.5	55.5
5	GPKT	66.5	86.2	60.1	29.8	50.2	58.6
顺序-2（D_1—D_5依次为DukeMTMC、MSMT17、Market-1501、CUHK-SYSU、CUHK03）							
1	DKP++	67.2	—	—	—	—	67.2
1	GPKT	67.0	—	—	—	—	67.0
2	DKP++	65.0	28.7	—	—	—	46.9
2	GPKT	63.6	30.2	—	—	—	46.9
3	DKP++	62.0	26.5	76.5	—	—	55.0
3	GPKT	62.1	28.2	78.1	—	—	56.1
4	DKP++	57.9	24.3	70.5	87.4	—	60.0
4	GPKT	61.9	26.9	70.6	88.9	—	62.1
5	DKP++	55.7	22.2	65.9	84.4	49.5	55.5
5	GPKT	58.8	25.6	69.4	86.8	52.0	58.5

注：行对应训练阶段t，列对应已见领域$\mathbf{D}_i$，单元格数值为$a_{t,i}$；“—”表示该领域在当前阶段尚未训练。

为进一步定位遗忘发生在哪些领域，图4给出了各旧领域的遗忘量$f_i = \max_{l\in\{i,\cdots,T-1\}} a_{l,i} - a_{T,i}$，即式(34)中求和的单项。图中横轴按训练顺序列出第1—4个已见领域，GPKT以红色实心柱表示，DKP++以蓝色斜线柱表示，遗忘量为mAP之差（百分点）。

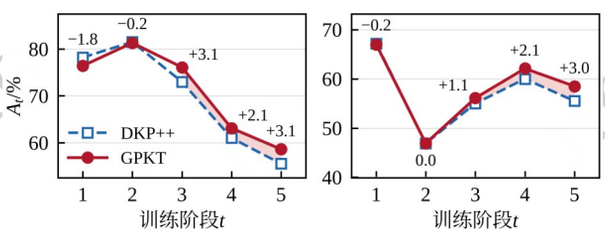
© 中国图象图形学报版权所有

表4　两种训练顺序下的遗忘与保持指标

Table 4　Forgetting and retention metrics under the two training orders

顺序	方法	A_T/% ↑	LA/% ↑	AF/pp ↓	BWT/pp ↑	RR/% ↑
顺序-1	DKP++	55.46	61.68	7.78	−7.78	87.38
顺序-1	GPKT	58.56	63.18	5.85	−5.78	90.53
顺序-2	DKP++	55.54	61.86	7.90	−7.90	85.74
顺序-2	GPKT	58.52	63.24	5.90	−5.90	89.76

注：↑表示越大越好，↓表示越小越好；A_T即表1、表2中的已见域平均mAP。为便于复核，本表各指标保留两位小数（表1、表2中保留一位小数）；AF以各旧领域在第i至第$T-1$阶段中的最佳性能为参照，BWT以其刚完成训练时的性能为参照；阶段平均性能、LA和RR的单位为%，AF和BWT为mAP之差，单位为百分点（percentage point，pp）。

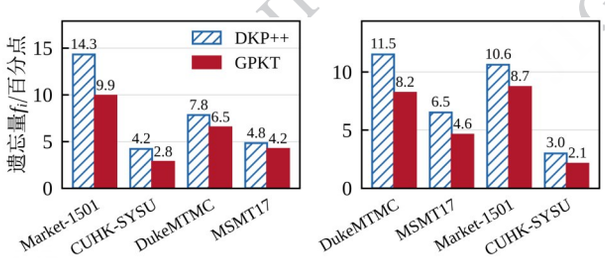


(a) 顺序-1　(b) 顺序-2

((a) Order-1; (b) Order-2)

图3　GPKT与DKP++的阶段平均性能A_t

Fig. 3　Stage-average performance A_t of GPKT and DKP++



(a) 顺序-1　(b) 顺序-2

((a) Order-1; (b) Order-2)

图4　GPKT与DKP++在各旧领域上的遗忘量f_i

Fig. 4　Forgetting f_i of GPKT and DKP++ on each old domain

两种顺序下，GPKT在全部旧领域上的遗忘量均小于DKP++，且最早训练的领域差距最大：顺序-1中Market-1501的遗忘量由14.3个百分点降至9.9个百分点，顺序-2中DukeMTMC的遗忘量由11.5个百分点降至8.2个百分点。这表明GPKT在各旧领域上均减轻了遗忘，且对经历后续更新次数最多的最早训练领域改善最为明显；结合式(37)的分解，这一保持收益与前述学习精度的提升共同构成了GPKT的最终性能优势。

2.4　未见域逐数据集结果与偏移类型分析

为避免平均值掩盖个别未见域的差异，7个未见域的逐数据集mAP/R1结果分别列于表5和表6，并在表7中将各数据集的主导分布偏移类型与GPKT相对DKP++的性能增益并列给出。

需要指出的是，7个未见域并非同质测试集，其主导分布偏移来源存在明显差异：CUHK01主要体现正/侧视角与姿态变化，CUHK02包含多组摄像头对带来的视角与场景变化，VIPeR具有低分辨率并伴随显著视角和光照变化，PRID突出跨摄像头的光照、背景与相机特性差异，i-LIDS主要受拥挤机场场景中的遮挡与背景杂波影响，GRID同时包含多摄像头跨视图变化和大量干扰身份，SenseReID则来自独立真实监控环境并采用检测器生成行人框，体现跨场景、跨摄像头及检测框差异。考虑到同一数据集可能同时包含多种因素，本文将上述类别视为主导偏移而非互斥标签，并在表7中结合逐数据集结果描述GPKT在不同数据集上的增益差异。

从表7结果可以看出，顺序-1下GPKT在7/7个未见域上均取得mAP和R1正增益，但不同域的增益幅度并不一致：GRID的mAP提升最高，为+4.9个百分点，VIPeR紧随其后为+4.8个百分点，二者的R1增益分别为+3.6和+4.0个百分点；i-LIDS为+3.9/+3.1个百分点，而SenseReID仅为+0.3/+1.1个百分点。顺序-2下，6/7个未见域同时取得正增益，GRID、VIPeR和i-LIDS的mAP/R1增益分别为+5.4/+3.8、+5.2/+4.1和+4.4/+3.5个百分点，SenseReID则为−0.4/−0.2个百分点。两种顺序的未见域平均提升分别为+3.2/+2.6和+3.3/+2.6个百分点。

两种训练顺序下的高收益域总体集中在VIPeR、i-LIDS和GRID，但具体增益幅度与排序随训练顺序发生变化；同时，SenseReID在顺序-1中保持小幅正增益、在顺序-2中出现轻微下降。上述结果仅说明逐域增益会随训练顺序发生变化。需要强调的是，表7中每种主导偏移类型通常仅对应一个代表数据集，且每个数据集同时包含视角、光照、分辨率、遮挡等多种因素，因此该表只能呈现增益与主导偏移类型之间的描述性对应，不足以据此判断某类偏移与GPKT特定模块之间的因果联系。在这一限

© 中国图象图形学报版权所有

表5　顺序-1下7个未见域的逐数据集性能(mAP/R1,%)

Table 5　Per-dataset performance on the seven unseen domains under Order-1 (mAP/R1, %)

方法	CUHK01	CUHK02	VIPeR	PRID	i-LIDS	GRID	SenseReID	未见域平均
PRD	31.6/24.3	38.3/31.8	8.8/6.9	29.3/24.5	20.6/14.1	9.0/6.7	23.4/15.6	23.0/17.7
PRAKA	56.8/49.5	61.8/55.1	37.9/32.6	52.2/48.0	44.6/38.0	32.0/28.2	48.6/39.8	47.7/41.6
FCS	62.8/53.3	66.8/57.1	40.2/35.3	58.0/49.0	46.5/40.5	36.8/31.3	53.6/42.9	52.1/44.2
ConRFL	65.7/59.9	73.2/65.6	45.3/42.6	61.6/60.0	54.3/47.3	43.1/38.8	58.6/51.9	57.4/52.3
MEGE	56.7/51.5	63.5/57.8	35.6/34.1	53.7/50.4	43.4/39.4	31.5/32.0	49.5/42.8	47.7/44.0
CKP	69.3/59.8	72.9/65.6	47.3/39.7	61.9/56.9	54.5/47.6	42.1/37.5	58.0/49.9	58.0/51.0
LSTKC	67.5/58.0	72.3/64.5	43.6/39.8	62.9/55.7	52.6/46.1	42.1/36.0	58.0/49.2	57.0/49.9
C^2R	60.4/51.4	62.6/55.2	37.6/32.0	51.9/47.6	45.7/37.2	32.4/28.1	50.3/41.8	48.7/41.9
DKP++	74.6/67.2	81.5/71.3	54.7/49.8	69.8/63.8	62.6/54.4	50.4/45.6	66.3/57.4	65.7/58.5
GPKT	76.8/68.9	85.1/74.2	59.5/53.8	72.5/65.6	66.5/57.5	55.3/49.2	66.6/58.5	68.9/61.1
GPKT-ViT	77.8/69.8	87.0/75.9	62.0/55.9	73.8/66.7	68.6/59.3	58.1/51.4	67.6/59.2	70.7/62.6

表6　顺序-2下7个未见域的逐数据集性能(mAP/R1,%)

Table 6　Per-dataset performance on the seven unseen domains under Order-2 (mAP/R1, %)

方法	CUHK01	CUHK02	VIPeR	PRID	i-LIDS	GRID	SenseReID	未见域平均
PRD	30.3/21.9	36.2/29.0	5.0/3.2	27.0/21.6	17.3/11.1	7.5/4.4	21.6/13.8	20.7/15.0
PRAKA	59.5/49.4	61.9/54.1	37.8/32.9	53.3/47.7	44.8/37.4	31.3/26.4	50.2/39.8	48.4/41.1
FCS	63.2/53.7	67.8/56.9	41.3/35.9	57.7/51.2	47.2/40.5	36.8/33.2	54.9/44.3	52.7/45.1
ConRFL	66.4/62.1	75.2/67.0	45.3/42.9	63.5/60.1	55.3/48.8	41.3/40.8	58.3/52.1	57.9/53.4
MEGE	52.8/48.4	59.7/54.6	32.4/30.4	50.1/48.1	40.4/37.9	29.5/27.5	45.2/40.8	44.3/41.1
CKP	68.2/57.9	72.4/64.8	44.5/39.9	62.1/56.4	53.8/45.4	41.8/35.5	57.6/50.1	57.2/50.0
LSTKC	67.1/56.9	73.8/64.4	45.5/39.0	62.0/55.5	53.2/46.4	42.9/36.0	58.7/49.0	57.6/49.6
C^2R	59.8/49.4	62.6/54.4	38.2/32.7	53.2/48.5	44.3/38.4	32.2/27.8	49.2/40.7	48.5/41.7
DKP++	75.1/65.7	79.8/72.0	52.6/48.4	70.1/64.0	60.3/53.1	50.7/45.4	65.7/56.7	64.9/57.9
GPKT	77.5/67.6	82.9/74.7	57.8/52.5	73.1/66.4	64.7/56.6	56.1/49.2	65.3/56.5	68.2/60.5
GPKT-ViT	79.1/68.7	85.2/76.2	60.5/54.4	74.3/67.2	66.8/58.3	58.6/51.2	66.2/57.3	70.1/61.9

注:GPKT-ViT采用ViT-B/16骨干,其余方法均采用ResNet-50骨干。

定下,可以观察到VIPeR、i-LIDS和GRID上的增益相对较大,而SenseReID上的改善较弱。

2.5　消融实验

为更直接地检验各功能单元之间的互补性,本文在顺序-1上采用完整的三因素组合消融。层次化分布原型记忆与关系蒸馏基本框架在所有变体中保持不变,仅考察原型迁移路径中的三个可独立控制单元:历史原型选择S、残差校准C和受治理关系蒸馏G。其中,G同时包含可靠性与风险感知的治理权重计算以及式(26)中的对数先验注入。之所以将二者合并,是因为治理权重只有注入样本—原型分配后才能实际影响关系教师;若只计算权重而不注入,其本身不会对优化产生独立作用。基于S、C、G三个单元,表8覆盖无增强基线、三个单模块、三个双模块以及完整S+C+G共8种组合,从而在相同训练协议下比较单独收益与联合收益。

表8显示,三个功能单元均具有独立贡献。相对基础关系蒸馏,S使已见域/未见域平均mAP由55.6%/65.4%提高到56.4%/66.3%,C提高到56.2%/66.0%,G提高到56.7%/66.6%;其中G的单

© 中国图象图形学报版权所有

表7　未见域主导偏移类型与GPKT相对DKP++的性能增益对应表(mAP/R1,%)

Table 7　Dominant shift types of unseen domains and gains of GPKT over DKP++ (mAP/R1, %)

未见域	主导偏移类型	顺序-1DKP++	顺序-1GPKT	顺序-1Δ	顺序-2DKP++	顺序-2GPKT	顺序-2Δ
CUHK01	视角/姿态偏移	74.6/67.2	76.8/68.9	+2.2/+1.7	75.1/65.7	77.5/67.6	+2.4/+1.9
CUHK02	多摄像头/视角偏移	81.5/71.3	85.1/74.2	+3.6/+2.9	79.8/72.0	82.9/74.7	+3.1/+2.7
VIPeR	低分辨率/成像质量偏移	54.7/49.8	59.5/53.8	+4.8/+4.0	52.6/48.4	57.8/52.5	+5.2/+4.1
PRID	光照/背景与相机特性偏移	69.8/63.8	72.5/65.6	+2.7/+1.8	70.1/64.0	73.1/66.4	+3.0/+2.4
i-LIDS	遮挡/背景杂波偏移	62.6/54.4	66.5/57.5	+3.9/+3.1	60.3/53.1	64.7/56.6	+4.4/+3.5
GRID	跨摄像头/干扰身份偏移	50.4/45.6	55.3/49.2	+4.9/+3.6	50.7/45.4	56.1/49.2	+5.4/+3.8
SenseReID	跨场景/检测框偏移	66.3/57.4	66.6/58.5	+0.3/+1.1	65.7/56.7	65.3/56.5	−0.4/−0.2
7域平均	**多类型综合偏移**	**65.7/58.5**	**68.9/61.1**	+3.2/+2.6	**64.9/57.9**	**68.2/60.5**	+3.3/+2.6

注:Δ=GPKT−DKP++,单位为百分点。偏移类型用于概括每个数据集最突出的域差异,并非互斥标签;同一数据集可能同时包含视角、光照、分辨率、遮挡等多种变化。由于每类主导偏移通常仅有一个代表数据集,本表仅用于描述性对应,不用于推断偏移类型与方法模块之间的因果关系。

表8　顺序-1上的组件组合消融与互补性分析

Table 8　Component-combination ablation and complementarity analysis on Order-1

组合设置	已见域平均		未见域平均	
	mAP	R1	mAP	R1
基础关系蒸馏(等权原型)	55.6	68.2	65.4	58.0
仅历史原型选择(S)	56.4	68.9	66.3	58.7
仅残差校准(C)	56.2	68.7	66.0	58.5
仅受治理关系蒸馏(G)	56.7	69.2	66.6	59.0
选择+校准(S+C)	57.2	70.0	67.3	60.0
选择+受治理关系蒸馏(S+G)	57.8	70.7	68.1	60.5
校准+受治理关系蒸馏(C+G)	57.7	70.6	68.0	60.4
完整GPKT(S+C+G)	**58.6**	**71.3**	**68.9**	**61.1**

独增益最大。双模块组合中,S+C达到57.2%/67.3%,S+G达到57.8%/68.1%,C+G达到57.7%/68.0%。S+C先减少无关原型并缓解空间失配,但候选原型仍以等权方式参与关系构造,因此整体性能低于包含G的两种组合;S+G与C+G的结果接近,则说明治理机制既能从更干净的候选集合中获益,也能从更准确的空间对齐中获益。

进一步以基础关系蒸馏为零点估计模块间的交互增益。S、C、G单独带来的已见域mAP增益分别为+0.8、+0.6和+1.1个百分点;若三者完全独立且仅作简单加和,则S+C、S+G和C+G的预期mAP分别为57.0%、57.5%和57.3%,而实际达到57.2%、57.8%和57.7%,对应额外交互增益为+0.2、+0.3和+0.4个百分点。未见域上的对应交互增益进一步达到+0.4、+0.6和+0.8个百分点。同样地,将三个单模块的独立增益简单相加,完整模型的已见域/未见域mAP预期为58.1%/68.1%,实际为58.6%/68.9%,分别高出0.5和0.8个百分点;R1也呈相同趋势,实际值71.3%/61.1%均比简单加和预期高0.9个百分点。另一方面,从任一双模块组合补入第三单元仍可获得稳定收益:相对S+C加入G后,已见域/未见域mAP进一步提升1.4/1.6个百分点;相对S+G加入C后提升0.8/0.8个百分点;相对C+G加入S后提升0.9/0.9个百分点。这些结果表明,S负责控制"迁移什么",C负责保证"原型是否对齐",G负责决定"迁移多少并如何进入关系教师",三者作用于迁移链条的不同环节,联合使用时形成正向互补,而非由某一单模块主导。

2.6　模型复杂度、训练开销与推理效率分析

为进一步评估GPKT的实际部署成本,本文统计了不同方法的参数量、浮点运算量(floating-point operations,FLOPs),并在相同输入尺寸与硬件环境下测试单张图像平均推理时间。所有对比方法均基于公开实现或依据原论文设置,并在本文统一实验协议下完成复现;参数量与FLOPs按统一配置统计,推理时间在NVIDIA A100图形处理器(graphics processing unit,GPU)上以256×128的输入分辨率实测。

© 中国图象图形学报版权所有

由于GPKT的原型选择、校准、治理与关系蒸馏均仅在训练阶段使用，推理阶段仍然只保留标准行人再识别特征提取器，因此不会引入额外的在线检索或原型匹配开销。不同方法的模型复杂度与推理效率结果如表9所示。

从表9可见，在ResNet-50骨干下，各方法的参数量约为25.8～29.4M、FLOPs约为4.18～4.42G，单张推理时间约为5.2～6.1ms，整体处于同一量级。GPKT的参数量、FLOPs和推理时间分别为26.9M、4.24G和5.5ms，低于ConRFL的28.2M、4.34G和5.8ms、C^{2}R的27.5M、4.29G和5.6ms，以及DKP++的27.8M、4.31G和5.7ms；与CKP的26.6M、4.22G和5.4ms也十分接近。这说明GPKT新增的原型治理与关系蒸馏主要发生在训练阶段，并未显著增加在线推理开销。GPKT-ViT采用ViT-B/16后模型规模与计算量明显增加，但更强的表征能力可带来进一步性能提升。整体而言，GPKT在精度提升与推理效率之间取得了较好的平衡。

表9　不同方法的模型复杂度与推理效率对比

Table 9　Comparison of model complexity and inference efficiency of different methods

方法	骨干网络	参数量/M	FLOPs/G	推理时间/ms
PRD	ResNet-50	25.8	4.18	5.2
PRAKA	ResNet-50	27.1	4.25	5.5
FCS	ResNet-50	26.4	4.21	5.3
MEGE	ResNet-50	28.6	4.36	5.8
ConRFL	ResNet-50	28.2	4.34	5.8
CKP	ResNet-50	26.6	4.22	5.4
LSTKC	ResNet-50	29.4	4.42	6.1
DKP++	ResNet-50	27.8	4.31	5.7
C^{2}R	ResNet-50	27.5	4.29	5.6
GPKT	ResNet-50	26.9	4.24	5.5
GPKT-ViT	ViT-B/16	86.6	11.34	13.2

除推理阶段外，GPKT的新增开销集中在训练阶段的原型选择、校准、治理与关系蒸馏。根据本文默认设置，GPKT采用ResNet-50骨干，输出特征维度$d = 2048$；每个小批量包含$B = 64$张图像、$P = 16$个身份，历史领域选择数$K = 3$，身份原型选择数$M = 12$。身份级检索对所选历史域的身份原型计算相似度，完成top-M检索后，仅12个历史原型进入校准、治理和关系蒸馏。将身份检索、边界风险计算、样本-原型分配和关系矩阵构造计入后，每个训练迭代的原型相关附加计算量按$2PdC_t + 2MPd + 2BMd + 2B^2M + 2B^2d$统计，其中$C_t$为$K = 3$约束下进入身份检索的历史身份数。记忆方面，每个历史身份保存2048维均值、2048维方差和样本支持计数，每个历史领域保存2048维全局均值与全局方差；32位浮点（32-bit floating point，FP32）精度下单个身份存储占用和单个领域描述符均约占16.00 KB。按本文五个已见域的标准训练划分，Market-1501、CUHK-SYSU、DukeMTMC、MSMT17和CUHK03分别包含751、5532、702、1041和767个训练身份。两种训练顺序下各阶段的原型相关附加计算与实际训练时间列于表10，累计训练身份数及对应的原型记忆存储占用列于表11。

表10按训练顺序分别列出各阶段的当前训练域及其训练图像数。第1阶段尚无历史记忆，不产生原型相关附加计算；自第2阶段起，附加计算随可检索历史身份数增加。由于K固定为3且M固定为12，第5阶段即使已有4个历史域，每个小批量仍只检索3个历史域，后续校准、治理和关系蒸馏也始终只处理12个原型。顺序-1的附加计算从第2阶段的0.070 GFLOPs/iteration增加到第5阶段的0.501 GFLOPs/iteration，顺序-2对应从0.067增加到0.501 GFLOPs/iteration。GPKT在ResNet-50下单张图像前向计算量为4.24 GFLOPs，批大小为64时单批主干前向计算量为271.36 GFLOPs；第5阶段原型相关附加计算占0.185%，表明训练侧新增计算受到K和M的显式约束，不随完整记忆池规模线性增长。

为进一步量化实际训练开销，本文在单张NVIDIA A100 GPU上记录GPKT在两种训练顺序下各阶段的实际wall-clock训练时间，结果亦列于表10。按照本文采用的不平衡训练划分（Pu等，2023），Market-1501、CUHK-SYSU、DukeMTMC、MSMT17和CUHK03的训练图像数分别为12 936、15 088、16 522、30 248和7 365，因此各阶段wall-clock时间主要受当前域数据规模影响，并不随历史阶段数单调增加。例如，顺序-1第4阶段的MSMT17包含30 248张训练图像，训练耗时19.8 min，而第5阶段的CUHK03仅包含7 365张训练图像，训练耗时4.9 min；此外，第1阶段训练80个epoch、其余阶段

© 中国图象图形学报版权所有

表10　GPKT各训练阶段的训练开销

Table 10　Training overhead of GPKT at each lifelong stage

阶段	训练域	训练图像数	历史域数	附加计算/G	相对前向/%	训练时间/min
顺序-1						
1	Market-1501	12 936	0	0.000	0.000	11.3
2	CUHK-SYSU	15 088	1	0.070	0.026	9.9
3	DukeMTMC	16 522	2	0.433	0.159	10.8
4	MSMT17	30 248	3	0.479	0.176	19.8
5	CUHK03	7 365	4	0.501	0.185	4.9
五阶段合计						56.7
顺序-2						
1	DukeMTMC	16 522	0	0.000	0.000	14.4
2	MSMT17	30 248	1	0.067	0.025	19.8
3	Market-1501	12 936	2	0.135	0.050	8.5
4	CUHK-SYSU	15 088	3	0.184	0.068	9.9
5	CUHK03	7 365	4	0.501	0.185	4.9
五阶段合计						57.5

注：训练图像数为按不平衡训练划分（Pu等，2023）得到的当前域训练图像数；第1阶段训练80个epoch，其余阶段训练60个epoch；附加计算为每次迭代的原型相关附加计算量（GFLOPs/iteration），相对前向为其占单批主干前向计算量（271.36 GFLOPs）的百分比。

训练60个epoch，这也使第1阶段的耗时相对较长。顺序-1与顺序-2的五阶段总训练时间分别为56.7 min和57.5 min，说明在完整五阶段学习过程中，两种训练顺序的总体训练成本接近。

表11　GPKT在两种训练顺序下的原型记忆增长

Table 11　Prototype memory growth of GPKT under the two training orders

阶段	顺序-1累计身份数	顺序-1记忆/MB	顺序-2累计身份数	顺序-2记忆/MB
1	751	11.75	702	10.99
2	6283	98.23	1743	27.27
3	6985	109.21	2494	39.03
4	8026	125.50	8026	125.50
5	8793	137.50	8793	137.50

原型记忆存储占用随累计历史身份数近似线性增长，如表11所示。由于每个历史身份仅保存固定维度的均值、方差与样本支持统计，每个历史领域仅保存固定维度的全局描述符，因此存储规模主要由累计历史身份数决定。两种顺序的中间阶段差异主要来自数据集在训练序列中的位置，例如包含5532个训练身份的CUHK-SYSU分别位于顺序-1的第2阶段和顺序-2的第4阶段；第5阶段两种顺序均累计8793个训练身份，最终原型记忆占用均为137.50 MB。历史记忆不包含历史行人图像。结合表9—表11可见，GPKT的额外成本主要表现为受固定候选规模约束的训练计算和随累计历史身份数线性增长的离线原型存储，而推理阶段不加载原型记忆，也不执行原型相关操作，因而保持零额外原型开销。

2.7　超参数敏感性分析

为分析关键超参数对模型性能的影响，本文在顺序-1上对GPKT进行单因素敏感性分析，并比较已见域平均与未见域平均上的mAP和R1。分析的超参数包括历史领域选择数K、候选历史身份原型数M、关系蒸馏温度τ、风险门控温度τ_r以及蒸馏损失权重λ；每次仅改变一个超参数，其余均固定为默认配置。需要强调的是，默认配置已按第2.1.5节所述验证集协议在任何测试集敏感性实验之前确定并锁定；本节及图5中的测试集结果均为锁定后的事后稳健性分析，仅用于描述参数扰动下的性能变

© 中国图象图形学报版权所有

化，不参与默认配置选择，也不反向修改任何训练设置。观察到，在当前较粗的候选网格中，默认配置恰好位于已见域与未见域平均mAP的最高点，而R1的最优点并不总与之重合。本文不将这一事后重合解释为“测试集最优配置”或调参收益：各超参数网格仅包含5个取值且间隔较粗，默认配置附近相邻取值的mAP变化大多在1个百分点以内，因此该结果只说明所考察网格内存在较平缓的高性能区域。表12—表16及图5的作用是报告锁定配置周围的稳健性，而非再次选择超参数。

由表12可见，K主要影响历史信息覆盖度与跨域噪声之间的平衡，但四项指标并不同步变化。K = 3时已见域/未见域mAP分别达到58.6%/68.9%；已见域R1在K = 4时达到71.6%，而未见域R1在K = 2时为61.2%，均与mAP的最优点不同。K继续增大到5时，已见域mAP下降至57.5%、未见域R1下降至60.0%，但未见域mAP仍为68.3%，说明不同检索指标对弱相关历史域的敏感程度存在差异。默认值K = 3由验证集协议确定；从测试结果看，该取值同样位于较优区间，且相邻取值（K = 2、4）下四项指标的变化均不超过1个百分点。

表12　不同历史领域选择数K下的敏感性分析（%）

Table 12　Sensitivity analysis under different numbers of selected historical domains K (%)

K	已见域平均mAP	已见域平均R1	未见域平均mAP	未见域平均R1
1	56.9	70.4	67.9	60.6
2	58.1	70.7	68.7	61.2
3	**58.6**	**71.3**	**68.9**	**61.1**
4	58.3	71.6	68.0	60.7
5	57.5	70.8	68.3	60.0

表13显示，身份原型数量M对整体排序质量和首位命中率的影响具有不同响应。M = 12时两项mAP达到58.6%和68.9%；已见域R1在M = 8时更高，为71.5%，未见域R1则在M = 16时达到61.3%。当M增大到20时，已见域R1明显回落至70.2%，但未见域mAP为68.5%，高于M = 16时的68.3%，表明候选规模增加并不会使所有指标按同一幅度退化。默认值M = 12由验证集协议确定；从测试结果看，M在8～16范围内四项指标的波动均不超过1个百分点，且M = 12的计算量适中。

表13　不同身份原型选择数M下的敏感性分析（%）

Table 13　Sensitivity analysis under different numbers of selected identity prototypes M (%)

M	已见域平均mAP	已见域平均R1	未见域平均mAP	未见域平均R1
4	57.8	71.0	68.2	60.1
8	58.2	71.5	68.7	60.4
12	**58.6**	**71.3**	**68.9**	**61.1**
16	58.4	70.9	68.3	61.3
20	57.9	70.2	68.5	60.8

表14　不同关系蒸馏温度τ下的敏感性分析（%）

Table 14　Sensitivity analysis under different relation distillation temperatures τ (%)

τ	已见域平均mAP	已见域平均R1	未见域平均mAP	未见域平均R1
0.03	57.2	69.8	68.5	60.7
0.05	58.0	70.6	68.8	60.1
0.07	58.5	71.6	68.1	60.9
0.10	**58.6**	**71.3**	**68.9**	**61.1**
0.15	57.7	70.9	67.6	60.5

表15　不同蒸馏损失权重λ下的敏感性分析（%）

Table 15　Sensitivity analysis under different distillation loss weights λ (%)

λ	已见域平均mAP	已见域平均R1	未见域平均mAP	未见域平均R1
0.3	57.5	70.0	68.4	61.0
0.5	58.4	70.9	68.0	60.5
0.7	**58.6**	**71.3**	**68.9**	**61.1**
1.0	58.2	71.7	68.6	60.2
1.5	57.6	70.5	67.4	59.8

表14给出了关系蒸馏温度τ的敏感性结果。较小τ会使样本—原型分配更尖锐，其影响首先体现在已见域R1：τ = 0.03时仅为69.8%，而τ = 0.07时升至71.6%。相比之下，未见域mAP在τ = 0.05时已达到68.8%，到τ = 0.07反而回落至68.1%；τ = 0.10时两项mAP分别达到58.6%和68.9%，未见域R1为61.1%。因此，τ对排序首位准确率与整体排序质量的作用方向并非简单同步，默认值τ = 0.10

© 中国图象图形学报版权所有

由验证集协议确定，其在测试集上的四项指标均处于较优水平。

表15表明，蒸馏损失权重λ体现了更明显的稳定性-可塑性权衡。$\lambda = 1.0$时已见域R1达到71.7%，高于默认配置，但未见域R1仅为60.2%；$\lambda = 0.3$时未见域R1仍有61.0%，而已见域mAP和R1分别下降至57.5%和70.0%。$\lambda = 0.5$时已见域mAP已达到58.4%，但未见域mAP仅为68.0%。这些差异说明增大蒸馏强度并不会使已见域与未见域、mAP与R1同步受益。默认值$\lambda = 0.7$由验证集协议事先确定；事后测试结果显示，该取值在两项mAP上处于高性能区间，两类域的R1也保持在较高水平。

表16　不同风险门控温度τ_r下的敏感性分析(%)

Table 16　Sensitivity analysis under different risk-gating temperatures τ_r (%)

τ_r	已见域平均mAP	已见域平均R1	未见域平均mAP	未见域平均R1
0.05	58.0	71.5	67.8	59.9
0.10	57.6	70.8	68.7	60.3
0.20	**58.6**	**71.3**	**68.9**	**61.1**
0.30	58.4	70.4	68.2	61.3
0.50	57.3	71.0	67.9	60.6

表16给出了风险门控温度τ_r的敏感性结果，其四项指标呈现更明显的分化。$\tau_r = 0.05$时已见域R1达到71.5%，但未见域R1降至59.9%；$\tau_r = 0.30$时未见域R1达到61.3%，同时已见域R1仅为70.4%。$\tau_r = 0.10$对未见域mAP较友好(68.7%)，但已见域mAP仅为57.6%。由验证集协议确定的默认$\tau_r = 0.20$时两项mAP达到58.6%和68.9%，并使两类域的R1保持在71.3%和61.1%的均衡水平。这表明风险门控温度过小或过大均可能导致部分指标下降。

上述单因素分析只能观察主效应，无法反映超参数之间的相互作用。为此，本文进一步在顺序-1上考察三组具有机制关联的二维组合：K与M共同决定候选历史原型规模，τ与λ共同决定关系监督的锐度和强度，τ_r与M共同影响风险抑制强度与候选规模的匹配。图5仅以已见域平均mAP和未见域平均mAP展示二维响应，以避免在热力图中重复堆叠R1指标；三组二维组合均采用5×5网格，并对每个配置独立运行5次后取平均。图5中(a)—(c)为已见域平均mAP，(d)—(f)为未见域平均mAP；(a)(d)、(b)(e)和(c)(f)分别对应历史领域选择数K与身份原型选择数M、关系蒸馏温度τ与蒸馏损失权重λ、风险门控温度τ_r与身份原型选择数M的耦合；单元格数值为对应超参数组合下的mAP(%)，红色方框表示默认配置，红色五角星表示该网格内的最优配置。当另一超参数固定为默认值时，相应行或列与表12—表16中的mAP单因素结果一致。

如图5所示，在本次粗网格的事后测试结果中，

组二维热力图的最高mAP位置基本与默认配置重合(默认配置已由验证集事先确定并锁定，见第2.1.5节)，即$(K, M) = (3, 12)$、$(\tau, \lambda) = (0.10, 0.70)$和$(\tau_r, M) = (0.20, 12)$，对应已见域平均mAP 58.6%、未见域平均mAP 68.9%。但三组表面的起伏形态明显不同：K-M网格的已见域/未见域波动范围分别为2.2/1.7个百分点，τ-λ为2.3/1.9个百分点，τ_r-M为1.8/1.6个百分点。各网格在默认点周围均存在连续高性能区域，但高值单元的分布并非简单复制同一条性能曲线。

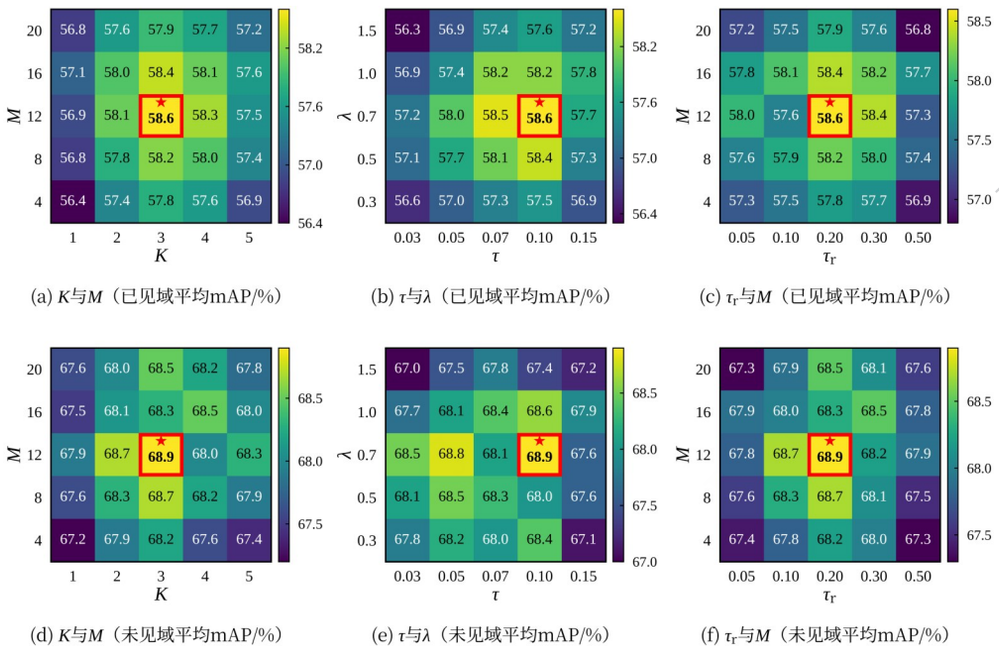
进一步观察局部结构，K-M网格主要表现为候选信息覆盖度与噪声规模的权衡：K较小时，适当增加M能够补充身份关系信息，而K较大时继续扩大M的收益有限。τ-λ网格中，已见域高值区域集中在$\tau = 0.07 \sim 0.10$、$\lambda = 0.5 \sim 1.0$附近，但未见域对部分组合的响应更不规则，例如$\tau = 0.07$在不同λ下并未保持同一排序，说明关系分布锐度与蒸馏强度存在域相关交互。τ_r-M网格则表现出风险抑制与候选规模之间的匹配关系：中等τ_r配合中等M最稳定，而过陡或过平滑的门控在不同M下造成的退化幅度并不一致。

总体而言，单因素与二维结果共享同一默认配置基准点，非默认设置下四项指标的局部最优位置、升降方向和变化幅度均存在差异。这表明mAP与R1、已见域与未见域对超参数的响应并不完全一致；在所考察的网格范围内，默认配置在多项指标之间取得了较稳定的综合表现。

2.8　多场景特征可视化与定量分析

为避免单一域转移场景可能带来的偶然性，并进一步考察不同训练阶段和训练顺序下特征表示变化的一致性，本文选取三组具有代表性的连续域转

© 中国图象图形学报版权所有



(a) K与M（已见域平均mAP/%） (b) τ与λ（已见域平均mAP/%） (c) τ_r与M（已见域平均mAP/%）

(d) K与M（未见域平均mAP/%） (e) τ与λ（未见域平均mAP/%） (f) τ_r与M（未见域平均mAP/%）

图5　超参数耦合敏感性热力图

Fig. 5　Heatmaps of hyperparameter coupling sensitivity

移场景进行特征分析，包括训练顺序-1中的Market-1501→CUHK-SYSU和DukeMTMC→MSMT17，以及训练顺序-2中的MSMT17→Market-1501。上述场景涵盖了不同数据集组合、不同连续学习阶段以及不同训练顺序，能够较为全面地反映模型在连续域学习过程中对历史知识的保持能力和对新领域的适应能力。针对每组域转移场景，本文分别比较MEGE、LSTKC、DKP++和GPKT所提取的身份特征。

如图6所示，整体采用三行四列的布局。其中，三行分别对应Market-1501→CUHK-SYSU、DukeMTMC→MSMT17和MSMT17→Market-1501三组连续域转移场景，四列依次对应MEGE、LSTKC、DKP++和GPKT四种方法。图中每种颜色对应一个身份：蓝色系（冷色调）标记表示前一领域$\mathbf{D}_{t-1}$中的身份，红色系（暖色调）标记表示当前领域$\mathbf{D}_t$中的身份；为便于黑白打印时区分，两类领域的身份还分别以圆形和三角形标记绘制。图6采用相同样本上的高维身份特征经t-SNE映射进行二维展示，用

于直观比较不同方法在类内紧凑性、类间可分性以及相邻领域混合程度上的相对差异。

考虑到仅依据二维散点分布进行判断可能具有一定主观性，本文同时给出轮廓系数（silhouette coefficient，SC）、局部类间/类内分离比（separation ratio，Sep）和最大均值差异（maximum mean discrepancy，MMD）三个量化指标。三项指标均基于降维前经L2归一化的高维特征计算。其中，SC采用标准轮廓系数，综合衡量样本对所属身份与其他身份的相对距离，其值越大表示聚类结构越清晰；Sep定义为各身份中心到最近竞争身份中心的平均距离，与各身份类内均方根半径平均值之比，其值越大表示局部身份分离越充分；MMD采用径向基函数（radial basis function，RBF）核衡量前一领域$\mathbf{D}_{t-1}$与当前领域$\mathbf{D}_t$的分布差异，核带宽由对应特征样本两两距离的中位数启发式确定，并以$\sqrt{\mathrm{MMD}^2}$的形式显示，其值越小表示跨域分布差异越弱。图6在各子图右上角统一采用“SC ↑ / Sep ↑ / MMD ↓”给出结果，二维t-SNE投影仅用于可视化。

从二维分布的整体结构看，四种方法在三个场景中仍呈现由MEGE、LSTKC到DKP++、GPKT逐步改善的总体趋势，但不同场景中的改善幅度并不相同。MEGE的身份特征整体较为分散，局部簇间混杂和领域相关聚集较明显；LSTKC在类内紧凑性与类间分离上有所改善，但不同领域身份仍存在区域性聚集；DKP++进一步形成更紧凑的身份簇和更清晰的局部边界，不过前后相邻领域之间仍保留一定分布差异。由此可以看出，方法层面的总体优劣趋势较稳定，但场景本身对聚类判别性与域间对齐的影响并非简单同步。

© 中国图象图形学报版权所有

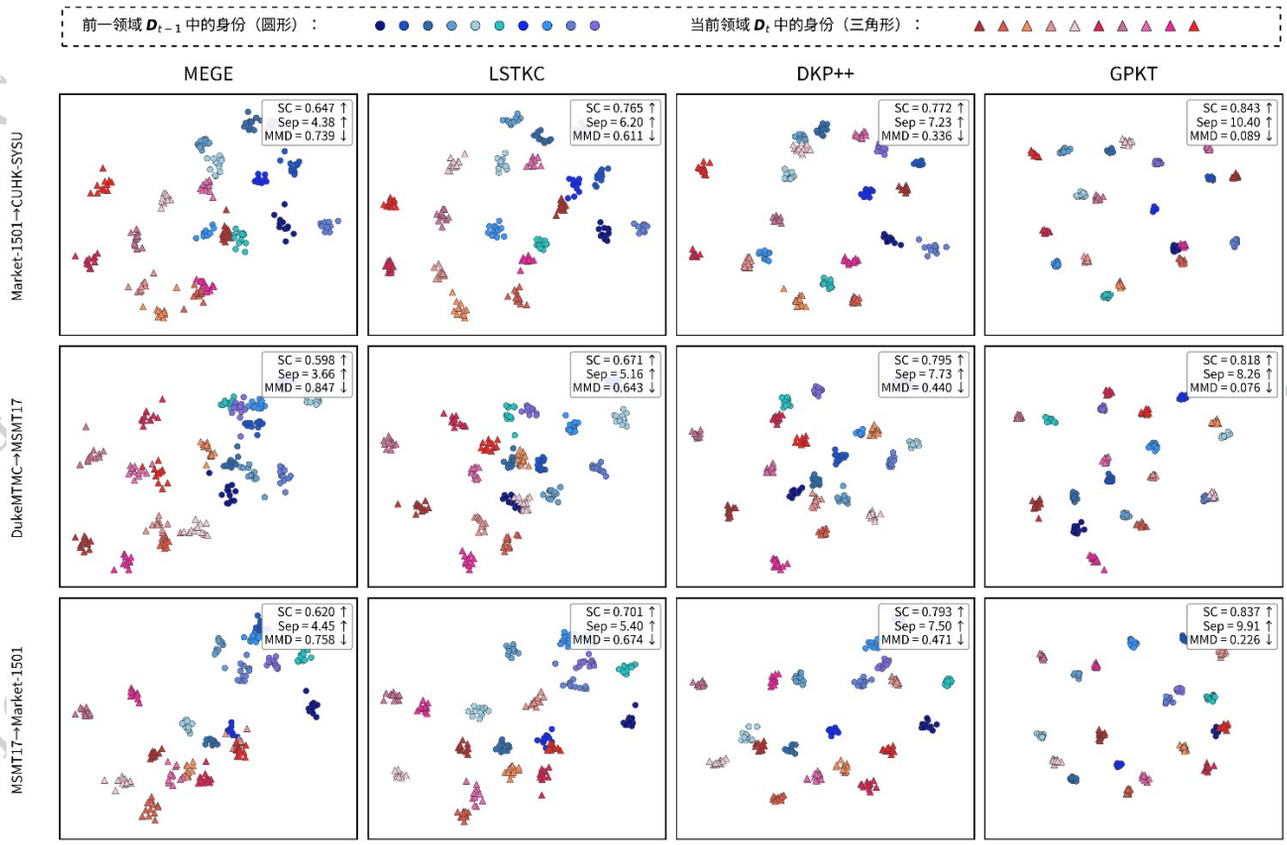


图6　多场景特征分布与量化分析

Fig. 6　Multi-scene feature distributions and quantitative analysis

相比之下，GPKT在三个连续域转移场景中均表现出更紧凑的身份结构和更弱的领域分隔，但其优势形式具有明显的场景依赖性。在Market-1501→CUHK-SYSU和MSMT17→Market-1501中，GPKT相对DKP++的类间分离优势更突出；而在DukeMTMC→MSMT17中，两者的SC和Sep差距相对较小，但GPKT的域间MMD下降更明显。

图6中的定量结果与上述场景差异相吻合。在Market-1501→CUHK-SYSU、DukeMTMC→MSMT17和MSMT17→Market-1501三组连续域转移场景中，GPKT的SC分别为0.843、0.818和0.837，Sep分别为10.40、8.26和9.91，MMD分别为0.089、0.076和0.226，在四种方法中均取得最佳综合指标表现。但三个指标对场景的排序并不相同：GPKT的SC和Sep均以Market-1501→CUHK-SYSU最高、MSMT17→Market-1501次之、DukeMTMC→MSMT17相对较低；MMD则在DukeMTMC→MSMT17中最低，其次为Market-1501→CUHK-SYSU，而MSMT17→Market-1501最高。该差异说明聚类判别性与域间分布差异反映的是不同统计性质，不应预期在不同跨域场景上呈现完全一致的相对变化。

进一步与三种对比方法中综合表现最优的DKP++比较。在上述三组域转移场景下，DKP++的SC分别为0.772、0.795和0.793，Sep分别为7.23、7.73和7.50，MMD分别为0.336、0.440和0.471。相比之下，GPKT的SC分别提高0.071、0.023和0.044，Sep分别提高3.17、0.53和2.41，MMD分别降低0.247、0.364和0.245。可以看到，三组场景不存在近似固定的增益倍率：Market-1501→CUHK-SYSU中的判别性提升最明显，DukeMTMC→MSMT17中的SC/Sep增幅较温和但MMD改善最大，MSMT17→Market-1501则保持较明显的Sep提升，同时其MMD仍高于另外两个场景。

综合t-SNE二维分布与对应量化结果可以看出，GPKT在三个场景中总体保持更好的身份聚类结构与跨域分布一致性，但不同指标的收益强度具有明显场景差异。这说明受治理原型迁移可能同时作用于类内紧凑性、局部类间分离和域间对齐

© 中国图象图形学报版权所有

3 结 论

本文提出了GPKT,一种面向无样本终身行人再识别的受治理原型知识迁移框架,将原型治理引入无样本终身行人再识别领域。GPKT将历史原型的利用方式从静态、无差别复用转变为选择性、校准化和可信度加权的知识迁移。具体而言,GPKT通过层次化分布原型记忆在不保留原始图像的前提下紧凑地存储身份级与领域级知识;通过相关性引导的原型选择与校准检索相关历史知识并缓解特征空间漂移;通过可靠性与风险感知的原型治理从样本支持、方差稳定性、领域相关性和边界风险四个维度评估原型可信度;通过受治理关系蒸馏将治理权重注入关系分布以保存历史身份关系结构。在12个行人再识别数据集上、按两种训练顺序与9种近期代表性方法的大量对比实验表明,GPKT在两种顺序下均优于综合平均性能最强的ResNet-50对比方法DKP++:顺序-1/顺序-2的已见域平均mAP/R1分别提升3.1/2.7和3.0/2.9个百分点,未见域平均分别提升3.2/2.6和3.3/2.6个百分点。逐阶段评估进一步表明,GPKT在两种顺序下的平均遗忘率较DKP++降低约2个百分点,学习精度提高1.4~1.5个百分点;其最终性能优势由旧领域保持与新领域适应共同贡献,二者大体相当,旧领域保持略占多数。这些结果表明,将相关性、可靠性与风险感知的原型治理与关系蒸馏相结合,能够在无需存储历史样本的前提下缓解灾难性遗忘并提升跨域泛化能力。

参考文献(References)

Asadi N, Davari M, Mudur S, Aljundi R and Belilovsky E. 2023. Prototype-sample relation distillation: towards replay-free continual learning//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA: PMLR: 1093-1106

Cui Z Y, Zhou J H, Wang X, Zhu M Y and Peng Y X. 2024. Learning continual compatible representation for re-indexing free lifelong person re-identification//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 16614-16623 [DOI: 10.1109/CVPR52733.2024.01572]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16×16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. Vienna, Austria: OpenReview.net

Feng Z X, Lai J H, Yuan Z, Huang Y L and Lai P J. 2025. Advancing universal person reidentification: a survey on the applications of large-scale, pretraining models for identifying individuals. Journal of Image and Graphics, 30(6): 1638-1660 (冯展祥, 赖剑煌, 袁藏, 黄宇立, 赖培杰. 2025. 走向通用行人重识别:预训练大模型技术在行人重识别的应用综述. 中国图象图形学报, 30(6): 1638-1660) [DOI: 10.11834/jig.240426]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network[EB/OL]. [2026-06-22]. https://arxiv.org/abs/1503.02531

Hu B, Xie G Q, Li L D, Li J, Yang J C, Lu W, et al. 2024. Progress of image retargeting quality evaluation: a survey. Journal of Image and Graphics, 29(1): 22-44 (胡波, 谢国庆, 李雷达, 李静, 杨嘉琛, 路文, 等. 2024. 图像重定向质量评价的研究进展. 中国图象图形学报, 29(1): 22-44) [DOI: 10.11834/jig.220722]

Huang J Z, Yu X H, An D, Wei Y G, Bai X, Zheng J, et al. 2023. Learning consistent region features for lifelong person re-identification. Pattern Recognition, 144: #109837 [DOI: 10.1016/j.patcog.2023.109837]

Jiang D and Ye M. 2023. Transformer network for cross-modal text-to-image person re-identification. Journal of Image and Graphics, 28(5): 1384-1395 (姜定, 叶茫. 2023. 面向跨模态文本到图像行人重识别的Transformer网络. 中国图象图形学报, 28(5): 1384-1395) [DOI: 10.11834/jig.220620]

Li Q W, Peng Y X and Zhou J H. 2024. FCS: feature calibration and separation for non-exemplar class incremental learning//Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 28495-28504 [DOI: 10.1109/CVPR52733.2024.02692]

Li Y Z, Fu L, Zhu L H, Luo Q D and Tu L. 2025. Multimodal large model-based method for generating visual Q&A data for electronic document images. Journal of Image and Graphics, 30(9): 3083-3096 (黎宇哲, 伏凌, 朱泠皞, 罗琪頔, 涂来. 2025. 多模态大模型面向电子文档视觉问答的数据生成. 中国图象图形学报, 30(9): 3083-3096) [DOI: 10.11834/jig.240610]

Li Z Z and Hoiem D. 2018. Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(12): 2935-2947 [DOI: 10.1109/TPAMI.2017.2773081]

Liu Z S, Guo Z W, Feng C L, Li W, Yu K, Hu J, et al. 2026. TP-LReID: lifelong person re-identification using text prompts. Pattern Recognition, 172: #112326 [DOI: 10.1016/j.patcog.2025.112326]

© 中国图象图形学报版权所有

Pu N, Chen W, Liu Y, Bakker E M and Lew M S. 2021. Lifelong person re-identification via adaptive knowledge accumulation//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 7901-7910 [DOI: 10.1109/CVPR46437.2021.00781]

Pu N, Zhong Z, Sebe N and Lew M S. 2023. A memorizing and generalizing framework for lifelong person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(11): 13567-13585 [DOI: 10.1109/TPAMI.2023.3297058]

Shi W X and Ye M. 2023. Prototype reminiscence and augmented asymmetric knowledge aggregation for non-exemplar class-incremental learning//Proceedings of 2023 IEEE/CVF International Conference on Computer Vision. Paris, France: IEEE: 1772-1781 [DOI: 10.1109/ICCV51070.2023.00170]

Sun B, Gao Y X, Zhuge W W and Wang Z X. 2023. Analysis of quality objective assessment metrics for visible and infrared image fusion. Journal of Image and Graphics, 28(1): 144-155 (孙彬, 高云翔, 诸葛吴为, 王梓萱. 2023. 可见光与红外图像融合质量评价指标分析. 中国图象图形学报, 28(1): 144-155) [DOI: 10.11834/jig.210719]

Wen W, Chen H and Zhang S L. 2026. Prompt-anchored vision-text distillation for lifelong person re-identification//Proceedings of 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Denver, USA: IEEE: 18503-18512

Wu G L and Gong S G. 2021. Generalising without forgetting for lifelong person re-identification//Proceedings of the 35th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press: 2889-2897 [DOI: 10.1609/aaai.v35i4.16395]

Xu K L, Zhang H Z, Li Y, Peng Y X and Zhou J H. 2024a. Mitigate catastrophic remembering via continual knowledge purification for noisy lifelong person re-identification//Proceedings of the 32nd ACM International Conference on Multimedia. Melbourne, Australia: ACM: 5790-5799 [DOI: 10.1145/3664647.3681235]

Xu K L, Zou X and Zhou J H. 2024b. LSTKC: long short-term knowledge consolidation for lifelong person re-identification//Proceedings of the 38th AAAI Conference on Artificial Intelligence. Vancouver, Canada: AAAI Press: 16202-16210 [DOI: 10.1609/aaai.v38i14.29554]

Yan J B, Tan Z W, Wu K C, Liu X L and Fang Y M. 2024. Viewport-independent and deformation-unaware no-reference omnidirectional image quality assessment. Journal of Image and Graphics, 29(12): 3699-3711 (鄢杰斌, 谭淄文, 吴康诚, 刘学林, 方玉明. 2024. 非视口依赖的抗畸变无参考全景图像质量评价. 中国图象图形学报, 29(12): 3699-3711) [DOI: 10.11834/jig.240188]

Yang L L, Lan L, Sun D T, Teng X, Ben X Y and Shen X B. 2023. Newly low-resolution pedestrian re-identification-relevant dataset and its benched method. Journal of Image and Graphics, 28(5): 1346-1359 (杨露露, 蓝龙, 孙冬婷, 滕霄, 贲晛烨, 沈肖波. 2023. 低分辨率行人重识别数据集及其基准方法. 中国图象图形学报, 28(5): 1346-1359) [DOI: 10.11834/jig.221082]

Yang S, Zhang Y F, Pu Y L and Yang H Y. 2025. View-aware feature learning for person re-identification. Journal of Image and Graphics, 30(1): 188-197 (杨山, 张永飞, 蒲养林, 杨航远. 2025. 面向行人再识别的朝向感知特征学习. 中国图象图形学报, 30(1): 188-197) [DOI: 10.11834/jig.240038]

Zhang Y F, Yang H Y, Zhang Y J, Dou Z P, Liao S C, Zheng W S, et al. 2023. Recent progress in person re-ID. Journal of Image and Graphics, 28(6): 1829-1862 (张永飞, 杨航远, 张雨佳, 豆朝鹏, 廖胜才, 郑伟诗, 等. 2023. 行人再识别技术研究进展. 中国图象图形学报, 28(6): 1829-1862) [DOI: 10.11834/jig.230022]

Zhou J H, Xu K L, Zhuo F, Zou X and Peng Y X. 2025. Distribution-aware knowledge aligning and prototyping for non-exemplar lifelong person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 47(12): 10932-10948 [DOI: 10.1109/TPAMI.2025.3597023]

作者简介

杨航远，男，博士研究生，主要研究方向为行人再识别和终身学习。E-mail：hyyang@buaa.edu.cn

张永飞，通信作者，男，教授，主要研究方向为计算机视觉。E-mail：yfzhang@buaa.edu.cn

陈思宇，女，研究员，主要研究方向为计算机视觉。E-mail：chensiyu6981@cetc.com.cn

马愈卓，男，博士研究生，主要研究方向为图像生成。E-mail：mayuzhuo@buaa.edu.cn

张雨佳，女，博士研究生，主要研究方向为计算机视觉和智能视觉监控。E-mail：yujia_zhang@buaa.edu.cn

李尚泽，男，博士研究生，主要研究方向为目标检索和模型压缩。E-mail：shangzeli@buaa.edu.cn

白彬沂，男，博士研究生，主要研究方向为视觉语言模型和多模态大模型。E-mail：bybai@buaa.edu.cn

郝晓辉，男，博士研究生，主要研究方向为计算机视觉。E-mail：by2306049@buaa.edu.cn

罗本，男，高级计算机视觉研究员，主要研究方向为智能视觉监控、行人再识别与野生动物监测。E-mail：sbakana@csir.co.za

© 中国图象图形学报版权所有