arXiv:2602.01012v1 [cs.CV] 1 Feb 2026

LocalScore：用于生物识别的局部密度感知相似度评分

Yiyang Su[1] Minchul Kim[1] JieZhu[1] Christopher Perry[1]Feng Liu[2] Anil Jain[1]
XiaomingLiu[1,3] 密歇根州立大学 [2] 德雷塞尔大学 [3] 北卡罗来纳大学教堂山分校{suyiyan1, kimminc2, zhujie4, perrych2}@msu.edu, fl397@drexel.edu, liuxm@cs.unc.edu

摘要

开放集生物特征识别面临探测主体可能未登记在图库中的挑战，因为传统生物识别系统难以检测这些非匹配探测样本。尽管多样本图库在实际部署中日益普遍，大多数现有方法仍将主体内变异压缩为单一全局表示，从而导致次优的决策边界和较差的开放集稳健性。为解决这一问题，我们提出 *LocalScore*，一种简单而有效的评分算法，它利用第 k个最近邻显式纳入图库特征分布的局部密度。*LocalScore* 架构无关、与损失函数无关，且计算开销可忽略不计，可作为现有生物识别系统的即插即用解决方案。跨多种模态的大量实验表明，*LocalScore* 在开放集检索（*FNIR@FPIR* 从53% 降至 40%）和验证（*TAR@FAR* 从 51% 提升至 74%）方面均持续取得显著提升。我们进一步提供理论分析与实证验证，依据数据集特征解释该方法在何时以及为何能获得最显著的增益。

1. 引言

生物特征识别通过面部特征等生理或行为特征识别个体 [23],步态模式 [49], 和人体形状 [68]。它是从访问控制到监控等各类应用不可或缺的一环 [46]。生物识别系统的核心是计算相似度分数，以评估探针与图库中已注册主体的匹配程度。验证用于判断探针是否与特定图库主体匹配，而检索则根据与探针的相似度对所有图库受试者进行排序。

在生物特征检索领域，早期研究主要关注闭集检索，即假定所有探测主体都存在于图库中（即，属于同源）。相比之下，开放集搜索代表了更贴近现实的推广，因为模型必须准确识别同源探针，同时拒绝非匹配探测样本

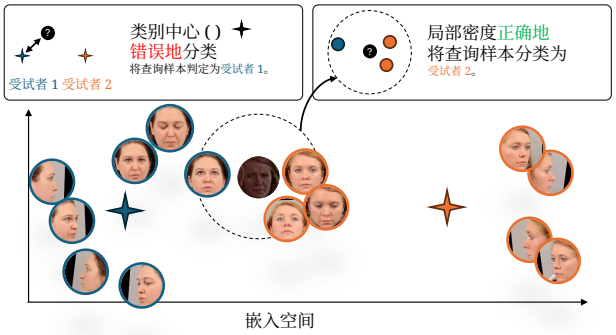


图 1. 局部密度在人脸识别中的作用示意图。每张人脸代表一个图库特征，不同的簇对应不同的视角。传统生物识别系统仅用一个类别中心（以星形表示）来表示每位受试者。然而，探测样本可能靠近该中心，却远离所有实际簇，从而在验证或开放集检索中导致错误接受。相比之下，图库特征分布的局部密度能够捕捉真实的主体内变异，并正确拒绝此类非匹配探测样本。

。这些要求使开放集检索本质上比仅涉及图库内识别的闭集检索更具挑战性。开放集评估已被认为是生物特征识别中最具挑战性的场景之一 [18]。尽管最先进水平（SoTA）的全身生物特征识别系统在闭集检索中的 Rank-20 准确率超过 90%，但当仅有 1% 的非匹配探测样本被错误识别为同源时，它仍无法正确识别约 55% 的同源探针 [38]。因此，本文聚焦于开放集检索与验证。

在实际部署中，为每个注册身份收集、处理和存储多个样本日益可行且常见，因此，多样本图库设置具有现实可行性和实际意义。因此，现代大型生物识别图库往往为每位受试者存储多个样本 [47]。例如，CCPG [36]平均每位图库主体包含 33 个视频。

尽管如此，大多数生物识别方法并未明确探讨如何利用多样本图库。现有方法通常较为简单，依赖于分数平均

在图库样本之间 [44]，或者假设每位受试者都可以用一个聚合特征嵌入来表示 [28,52]。这些模型从输入样本（例如，图像或视频）中提取生物特征，然后将其聚合为每位受试者的单个类别中心，忽略了多样本图库中的个体内差异。

天真地对图库特征嵌入或分数进行平均，无法捕捉该受试者的真实变化。如图 1（左上）所示，这类方法仅依赖于与受试者全局中心的相似度，而忽略了会在同一受试者内部形成不同簇的多种条件（例如，姿态、光照和视角）。因此，将多簇图库分布压缩为单一中心，会导致决策边界不理想，并削弱判别能力。这一局限性在开放集场景中尤为不利，因为要准确拒绝未见过的受试者，就需要对受试者内变化的完整范围进行建模。

为说明这一问题，图 1 展示了一个人脸识别示例，其中示例受试者的图库嵌入形成了多个与不同相机视角相关的明显簇。探针可能位于受试者 #1 的中心附近，却仍然远离其各个簇，从而导致将非匹配样本错误接受。相反，来自代表性不足簇的真实探测样本可能会被错误拒绝。这些失败案例凸显了仅依赖类别中心的局限性，并促使人们需要采用能够利用图库特征局部密度 的方法，以更准确地反映受试者内变化。

我们提出了 LocalScore，这是一种新颖的评分框架，在生物识别系统中通过局部结构而非全局平均来建模主体内变异。其核心思想是，k阶最近邻（k-NN）为捕捉受试者图库特征的局部几何结构提供了一种简单而有效的方法。如图 1所示，非同源探针可能靠近全局图库中心，却仍远离所有真实簇，从而导致基于传统中心的评分错误地接受该探针。LocalScore 通过使用 k-NN，评估探针与真实图库样本局部邻域的匹配程度，从而解决了这一问题。与最近邻距离较远的探针更有可能是冒名者，而真实探针通常距离最近邻更近。借助局部图库结构，LocalScore 显著提升了验证准确率和开放集鲁棒性。此外，LocalScore 易于部署，仅需一行代码即可实现。它是一个即插即用模块，与任何生物识别模型兼容，不受架构或损失函数影响，并且仅引入可忽略不计的计算开销。

我们在三种主要的生物特征模态上评估 LocalScore：人脸识别、步态识别和行人重识别，并采用多种模型架构，包括 ResNet [16], ViT [39, 64], 和 Swin Transformer [11]。这些模型使用多种损失函数进行训练，例如交叉熵变体 [27] 和三元组损失 [20]。在所有设置下，LocalScore 始终优于标准评分方法，在开放集性能方面取得显著提升（FNIR@FPIR 从 53% 降至 40%），验证准确率也有所提高（TAR@FAR 从 51% 升至 74%）。

此外，我们还对所提出的算法进行了全面的理论和实证分析，以考察其行为。从理论上讲，我们基于分数统计对使用和不使用 LocalScore 时的预期性能进行建模，并指出 LocalScore 在哪些条件下有望提升开放集生物特征识别和验证的效果。从实证角度看，我们通过真实世界数据集和蒙特卡洛模拟验证理论见解的可靠性，并考察 k 在 k-最近邻以及受试者内变化对性能的影响。我们发现，该算法在各种条件下都能持续提升性能，体现出良好的稳健性。

我们的贡献总结如下：

- 我们提出了 LocalScore，将每个探针周围图库特征分布的局部图库密度信息纳入生物识别分数。
- 我们验证了该方法在多种生物识别模型和任务中的优越性能，显著提升了识别准确率。
- 我们从理论层面、通过合成数据集上的模拟以及在真实世界数据集上，对传统距离度量与我们的算法进行了详细比较分析。

2. 相关工作

生物特征识别。 生物特征识别已在多种模态中得到广泛研究，例如人脸识别 [9, 30, 32,43, 66], 步态识别 [19, 21, 24, 50, 59, 63, 65], 行人重识别 [5, 8, 26, 35, 37, 42, 53, 57, 60, 67, 69], 以及模态组合 [31, 38, 40, 73]。大多数既有工作侧重于设计网络架构或损失函数，以获得更具判别性的表征。然而，评分通常依赖于受试者聚合后的图库中心，忽略了图库特征的内在结构。相比之下，我们提出了 LocalScore，利用图库特征的局部密度生成更可靠的相似度评分。此外，尽管许多现有方法侧重于验证或闭集检索，但只有少数方法 [17, 52] 明确针对更具挑战性的开放集场景。

开放集识别。 开放集识别 [3, 7, 33, 34,45, 56, 58] 旨在区分图像分类中的已知类别和未知类别。如 [52], 所述，已知类别对应训练集中的类别，而未知类别则未用于训练 [13]。相比之下，即使在闭集生物识别中，测试受试者也未用于训练，而同源探针是已在图库中登记的测试受试者 [22]。开放集生物特征识别与闭集生物识别的区别在于，非匹配受试者既不出现在训练集中，也不出现在图库中 [15, 17]。

尽管先前研究已探索将 k-最近邻用于分布外（OOD）检测 [54] 和开放集识别 [61]，k-最近邻分数本身并不具备足够的判别信息，无法用于生物特征识别。相比之下，我们的方法利用 k-最近邻对探针周围的局部特征密度进行建模，并将这一密度度量整合到生物特征评分中。

重排序。 重排序 [1, 2, 41, 48, 51, 55, 62, 70, 72]能够增强预训练行人重识别模型对探测样本生成的原始排序。值得注意的是，[72] 通过考察探针是否属于图库中最近邻的邻居来提升性能。虽然重排序可应用于各种生物特征模态，但要有效运行，它依赖大量探测样本或查询样本。然而，在现实应用中，探测样本往往是连续接收的，其分布未知。因此，当只有一个探测样本可用时，重排序的理论基础和性能都会受到影响。相比之下，我们的方法独立处理每个探测样本，无需依赖其他探测样本。此外，由于方法简单，我们的方法比重排序的计算效率更高。

```
定义 局部_分数(模型, x, G_S, G_M, k):
    """
    参数:
    - model: 预训练生物特征模型
    - x: 探测输入
    - G_S: (N, D) 按受试者划分的图库数据库
    - G_M: (N_m, D) 按媒体划分的图库数据库
    - k: KNN 中的"k"

    返回:
    - (D,) 更新后的分数向量
    """

    # (N_g, ); 与 G_S 的相似度分数
    相似度_分数 = 相似度(模型(x), G_S)

    # (1,); G_M 中的第 k 近邻
    knn_score = sim(model(x) , G_M).sort()[ , -k]

    # (N_g, ); 输出相似度分数
    相似度_分数 += knn_分数 * (相似度_分数 ==
        相似度_评分。最大值())

    返回 相似度分数_
```

算法 1. LocalScore 的 Python 风格伪代码。

3. 方法

3.1. 问题设定

设有一个包含 N 个对象的图库，每个对象都有多个样本（例如，图像或视频）。对于第 i 个对象，其生物特征为 $\mathbf{g}_i^1$，……，$\mathbf{g}_i^{m_i} \in \mathbb{R}^d$，完整图库为$\mathcal{G} = \{\mathbf{g}_i^j \mid 1 \le i \le N,\ 1 \le j \le m_i\}$。给定查询样本 $\mathbf{q} \in \mathbb{R}^d$，目标是计算 $\mathbf{s} = f(\mathbf{q}, \mathcal{G})$，其中每个 $\mathbf{s}_i$ 衡量$\mathbf{q}$ 与对象 i 之间的相似度。f 的设计直接影响验证和开放集检索的性能。

一种常见策略 [52]使用单个聚合特征表示每个受试者：

$$\mathbf{s}_i = \left\langle \frac{\mathbf{q}}{\|\mathbf{q}\|}, \frac{1}{m_i}\sum_{j=1}^{m_i} \frac{\mathbf{g}_i^j}{\|\mathbf{g}_i^j\|} \right\rangle. \tag{1}$$

EVM [17] 和 CAFace [28] 等方法都遵循这一原则，仅在聚合细节上有所不同。然而，它们都将受试者内变化压缩为单个原型，丢弃了特征分布的结构。

另一种方法是对逐样本相似度取平均：

$$\mathbf{s}_i' = \frac{1}{m_i}\sum_{j=1}^{m_i} \left\langle \frac{\mathbf{q}}{\|\mathbf{q}\|}, \frac{\mathbf{g}_i^j}{\|\mathbf{g}_i^j\|} \right\rangle, \tag{2}$$

这种方法保留了所有样本，但将它们一视同仁，忽略了跨样本相关性或局部结构。

现有的评分函数为每个受试者独立计算 f，而忽略了图库分布。具体而言，传统相似度假设受试者的特征形成一个单一紧密簇，而真实世界的数据往往包含对应于不同视角、光照或其他条件的多个不同区域。因此，靠近稀疏区域的探测样本可能看起来接近全局均值，但与任何真实区域都不对齐，从而导致错误接受或拒绝。忽视这种局部结构会限制传统评分函数的可靠性。

3.2. LocalScore

由于图库特征自然反映了受试者内部的变化，我们在 $\mathcal{G}$ 评分时纳入所有样本，以捕捉每位受试者的特征模式。我们不对所有 m_i 样本的图库特征或成对相似度取平均，而是关注每位受试者特征空间的局部几何结构。我们在图库样本中找到探针 q 的第 k 近邻，并计算

$$s^{(k)} = \underset{i,j}{\text{top-k}} \left(\left\langle \frac{\mathbf{q}}{\|\mathbf{q}\|}, \frac{\mathbf{g}_i^j}{\left\|\mathbf{g}_i^j\right\|} \right\rangle \right), \tag{3}$$

其中，top-k(·) 返回第 k 大的相似度，而 $\mathbf{g}_i^j$ 表示受试者 i 的第 j 个特征。这衡量了探针与图库分布中最相关局部区域的契合程度。真实探测样本通常会得到较大的 $s^{(k)}$，因为它们聚集在正确受试者图库的密集区域内；而冒名者通常位于这些模式之外，因此得分较低。

为在保留身份信息的同时纳入局部几何结构，我们将 k- 最近邻分数与标准的逐受试者分数 $\mathbf{s}$ 融合。对 $\mathbf{s}_i$ 和 $\mathbf{s}^{(k)}$进行简单平均会降低性能，因为$\mathbf{s}^{(k)}$ 缺乏明确的身份监

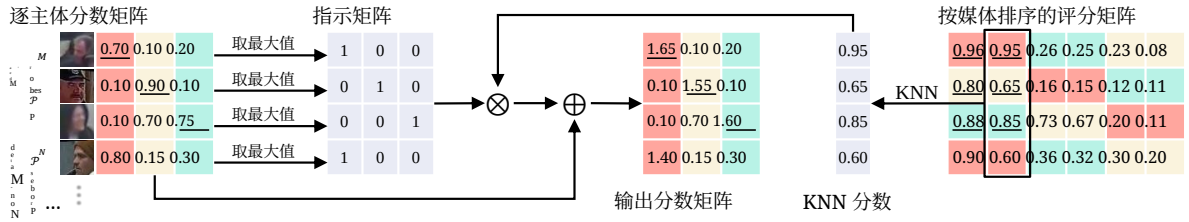


图 2. LocalScore 的一个简单示例，其中真实分数以粗体和下划线标出。不同颜色代表不同的图库受试者。对于每个探针，根据逐媒体分数矩阵计算 k-最近邻分数，并将其加到每主体最大分数上，以生成输出分数矩阵。例如，对于第一个探针，将 k-最近邻分数 0.95 加到最高主体得分 0.7 上，而所有其他分数保持不变。在原始分数矩阵中，只有一个真实分数超过了 0.8 的每探针最大非同源分数。然而，应用 LocalScore 后，所有真实分数都超过了更新后的 1.4 的每探针最大非同源分数。

督，导致非匹配分数虚高，并缩小真实匹配与冒名者之间的差距。例如，我们发现，这种简单的融合策略会使 KPRPE 在 [29] BRIAR数据集 [6] 上的 TAR@0.1%FAR 从 60% 降至 30%。因此，我们仅对每个探针的 k最高分主体应用 -NN 调整。

例如，如图 2 中第三个同源探针所示，在进行朴素均值融合后，它与第一个受试者的相似度为，即 $(0.70 + 0.85)/2 = 0.775$。该值等于第二个探针与其真实匹配对象（第二个图库主体）之间的相似度，即 $(0.90 + 0.65)/2 = 0.775$。在此示例中，第一个分数是非匹配分数，而第二个是匹配分数。因此，由于匹配分数与非匹配分数之间的差距消失，导致被计为真阳性的探针减少，验证性能因此下降。因此，需要一种更复杂的算法。

为解决这一问题，令$\mathbf{i} \in \{0,1\}^N$ 为一个指示向量：若$\mathbf{i}_i = 1$ 满足$\mathbf{s}_i = \max_j \mathbf{s}_j$，则取该值，否则取 0。最终的融合分数$\mathbf{t} \in \mathbb{R}^N$ 计算如下：

$$\mathbf{t} = \mathbf{s} + \mathbf{i} \cdot \mathbf{s}^{(k)}. \tag{4}$$

这种选择性融合仅提高最可能受试者的相似度，从而避免分数整体膨胀。LocalScore 如算法 1 所示。请注意，它仅利用局部图库特征分布，不使用图库样本的标签。

与朴素融合策略相比，LocalScore仅提高每个探针中得分最高的受试者的相似度分数，以避免非匹配分数的普遍抬升。在图2所示的简单示例中，这种选择性增量使第三个探针与第二个图库主体之间的相似度保持在0.70，远低于任何真实匹配分数。因此，非匹配分数不会与真实匹配重叠。通过不抬高非最高排名对象的分数，LocalScore在冒充者与真实分布之间保持了更宽的分离，从而提升验证性能。

LocalScore 无需训练、高效且与现有生物识别模型完全兼容。它不假设某位受试者的图库特征构成单个紧密簇，而是对其局部结构进行建模，从而提升验证和开放集识别能力，且无需更改网络架构或训练目标。

3.3. 图库聚类

虽然纳入全部图库样本可通过捕捉多样的受试者内变异来增强稳健性，但也会增加样本数量较多的对象所需的内存与计算。为缓解这一问题，我们可选择将每位受试者的图库特征聚类为一组紧凑的代表性原型，在降低成本的同时保留局部结构。

对于第 i个具有图库特征 $\{\mathbf{g}_i^1, \ldots\ldots, \mathbf{g}_i^{m_i}\}$的受试者，我们将其聚类为 $C_i \ll m_i$ 个中心 $\{\mathbf{c}_i^1, \ldots\ldots, \mathbf{c}_i^{C_i}\}$，其中每个中心为

$$\mathbf{c}_i^k = \frac{1}{|\mathcal{C}_i^k|} \sum_{\mathbf{g}_i^j \in \mathcal{C}_i^k} \mathbf{g}_i^j, \tag{5}$$

且 $\mathcal{C}_i^k$ 是第 k 个簇中的特征集合。最终得到的紧凑图库为 $\tilde{\mathcal{G}} = \{\mathbf{c}_i^k\}_{i,\ k}$。

在推理过程中，所有相似度计算（包括k-最近邻搜索和融合）都在聚类中心上进行。每个受试者的 k-最近邻得分变为

$$\mathbf{s}_i^{(k)} = \underset{j}{\text{top-k}} \left(\left\langle \frac{\mathbf{q}}{\|\mathbf{q}\|}, \frac{\mathbf{c}_i^j}{\|\mathbf{c}_i^j\|} \right\rangle \right), \tag{6}$$

其中，每个 $\mathbf{c}_i^j$ 捕获了受试者 i 分布的一个局部模态。这种紧凑表示以信息丰富的原型取代冗余样本，在准确率与效率之间实现了良好平衡。

3.4. 理论分析

我们建立了一个理论框架，用于定义 LocalScore 的适用边界，从而为其能够提升验证和开放集搜索性能的场景提供界限。具体而言，定理 1 提供了一个判据，用于评估k-NN 分数是否能够在同源探针与非同源分数之间提供足够的区分度，从而同时改善 FNIR@FPIR 和 TAR@FAR。

假设每个对象的真实分数服从正态分布 $\mathcal{N}(\mu_1, \sigma_1)$。每个对象的冒名者分数服从分布 $\mathcal{N}(\mu_2, \sigma_2)$。同源探针的 k-NN 分数服从分布 $\mathcal{N}(\mu_3, \sigma_3)$，而非同源探针的 k-NN 分数服从分布 $\mathcal{N}(\mu_4, \sigma_4)$。此外，设 N_1为同源探针的数量，N_2 为非同源探针的数量，M 为图库受试者的数量，N_3 为冒名者分数的总数，r_1 和 r_2 为可容忍的每个探针样本的最大非同源分数数量（例如，1% $\times N_2$）和非匹配分数数量（例如，0.1% $\times N_3$）。

定理1。 算法 *1*在期望阈值下提升期望开放集 *FNIR@FPIR*，当

$$\frac{\mu_2 + \mu_4 + \delta\sqrt{\sigma_2^2 + \sigma_4^2} - (\mu_1 + \mu_3)}{\sqrt{\sigma_1^2 + \sigma_3^2}} < \frac{(\mu_2 + \delta\sigma_2) - \mu_1}{\sigma_1}, \tag{7}$$

其中 $\delta = -\ln(-\ln(1 - r_1/N_2))$并在期望阈值下提升验证 *TAR@FAR*，当

$$\frac{\mu_2 + \mu_4 + \Phi^{-1}\left(\frac{N_3 - r_2 - \alpha}{N_3 - 2\alpha + 1}\right)\sqrt{\sigma_2^2 + \sigma_4^2} - (\mu_1 + \mu_3)}{\sqrt{\sigma_1^2 + \sigma_3^2}} < \frac{\mu_2 + \Phi^{-1}\left(\frac{N_3 - r_2 - \alpha}{N_3 - 2\alpha + 1}\right)\sigma_2 - \mu_1}{\sigma_1}, \tag{8}$$

其中 $\Phi^{-1}(\cdot)$ 是正态分布累积分布函数的反函数，$\alpha = \pi/8$ 是一个常数。

定理的证明请参见补充材料。1

4. 实验

评估协议。在条件允许时，我们遵循 [52] 的评估协议（训练与测试、图库与探针，以及匹配与非匹配划分）。对于较新的数据集，我们以相同方式随机生成评估协议。

评估指标。我们从两个方面评估生物识别模型：验证和开放集检索。对于验证，我们报告在.1%和1%阈值下的TAR@FAR。对于开放集搜索，对于较大的数据集，我们报告在.1%和1%阈值下的FNIR@FPIR；对于较小的数据集，则报告在1%和5%阈值下的FNIR@FPIR。鉴于FNIR@FPIR [52], 存在固有的变异性，我们根据50次运行的结果提供其均值和95%置信区间。

数据集。 除 IJB-S、Gait3D 和 CCVID [52],外，我们还在 CCPG [36] 上评估基于 RGB 的步态识别，并在 BRIAR [6] 上评估多模态融合。CCPG 面向换装步态识别和行人重识别。该数据集包含 100 名训练受试者和 100 名测试受试者，分别有 8,388 个和 8,178 个序列。CCPG 提供人脸和鞋子区域被灰化的 RGB 图像，使基于 RGB 的步态识别方法（如 BigGait [64], ）能够应用其中，且其性能优于基于轮廓的方法。BRIAR 关注跨模态的远距离及基于无人机的生物特征识别，涉及人脸、步态和行人重识别。其训练集包含 995 名受试者；我们遵循官方协议 v6.1.0, 使用 1,280 名图库受试者和 6,813 个探针序列，评估 LocalScore 在全身人体识别中的性能。

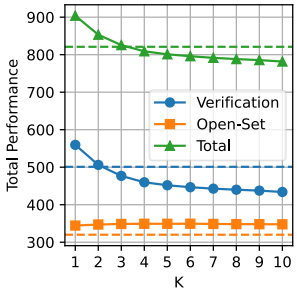


图 3。 k 对整体性能的影响。总性能为 TPIR@FPIR（即，1−FNIR@FPIR）、TAR@FAR 和 Rank-1准确率之和。虚线：不使用 LocalScore。

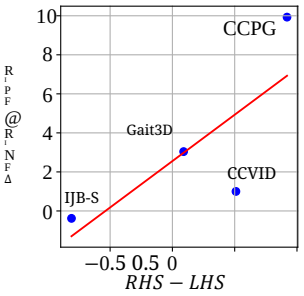


图 4 LocalScore 在不同数据集上带来的性能提升与式（7）右侧与左侧理论差值之间的关系。红线：线性回归。

4.1. 消融研究

的影响 k。我们分析了三种模态下验证和开放集识别性能对 k 的敏感性：基于 BRIAR 的人脸识别 [6] 与 IJB-S [25], 、基于 Gait3D 的步态识别 [71] 与 SwinGait [11], , 以及基于 CCVID 和 CAL 的行人重识别 [16]。图3 展示了验证性能与开放集搜索性能之和。开放集性能在不同 k 取值下仍然稳健，表明 LocalScore 具有良好的泛化能力。对于验证任务，性能会随着 k 增大而下降。总体而言，较小的 k 取值能带来更好的性能，尤其是在图库规模较小的数据集上（例如，Gait3D 平均每个主体仅有 5.4 个序列）。较大的 k取值往往会降低验证性能，因为真实探测样本更有可能包含错误图库身份的邻居。因此，我们在实验中选择 k=1 。

图库聚类的影响。 为考察存储与性能之间的权衡，我们评估了在评分前对每位受试者的图库进行聚类的效果。具体而言，图库使用 k-均值聚类为 C_i 个簇，并以各簇的质心作为代表。

如表1所示，开放集 FNIR@FPIR 会随着 C_i 的增加而改善。这表明，更全面地覆盖受试者的受试者内变化，可以减少开放集场景中的误拒绝。相比之下，验证（TAR@FAR）在不同簇规模下保持稳定，这说明只需少量质心就足以完成真实匹配识别。

模型	数据集	C_i	% 画廊	Run-time	FNIR@FPIR ↓		TAR@FAR ↑	
					0.1%	1%	0.1%	1%
SwinGait [11]	Gait3D†	N/A	< 1	124	81.84	66.09	57.81	78.95
		5	53.7	246	83.50	69.10	**63.65**	78.95
		20	92.3	319	81.01	65.43	**63.65**	**79.08**
		∞	100	338	**78.80**	**62.98**	**63.65**	**79.08**
BigGait [64]	CCPG [36]	N/A	< 1	29	46.95	34.53	82.46	95.64
		5	14.2	60	41.04	29.52	92.25	97.00
		20	56.8	175	39.23	27.34	92.89	97.00
		∞	100	301	**37.04**	**25.34**	**92.95**	**97.01**

表1. 不同 C_i下的开放集 FNIR@FPIR 和验证 TAR@FAR。% 画廊：聚类后图库大小与完整图库大小的比值。运行时间以毫秒（ms）为单位。†FNIR@FPIR 报告了阈值为 1% 和 5% 时的结果。

不同数据集对聚类的敏感度各不相同。Gait3D [71] 中每位受试者的样本数量较少且变化较大（平均值 ≈ 5），因此聚类带来的收益有限。相比之下，在受控条件下，CCPG [36],中每位受试者拥有更多样本（平均值 ≈ 35），使用不到 50% 的样本即可达到接近完整的性能。因此，可以压缩冗余的图库特征，同时将准确率损失降至最低。

在后续实验中，我们使用完整图库报告最佳性能。

4.2 结果

与最先进水平的比较。我们使用涵盖多种生物特征模态、数据集和图库聚合方法的最先进水平模型评估 LocalScore，包括用于人脸的 CAFace [28]，以及通过对图库特征拟合威布尔分布的 EVM [17], 。如表2所示，在所有设置下，LocalScore 均能稳定提升原始分数。例如，在 CCPG 上，它将 BigGait 的 TAR@0.1%FAR 从 82.46 % 提升至 92.95%（提高 10.49%），并将 FNIR@0.1% FPIR 从 46.95% 降至 37.04（降低 9.93%），且不会降低任何模型的性能。与 EVM 结合时，LocalScore 还能进一步提升验证准确率，例如将 Gait3D 上的 TAR @0.1%FAR 从 69.68% 提升至 79.84%（提高 10.16%），同时保持较低的推理成本，且只需一个超参数。

理论预测。 定理 1提供了一种理论工具，可用于预测 LocalScore 在现实场景中的性能。为验证理论预测的准确性，我们比较了 LocalScore 带来的性能提升与式（7）右侧（RHS）和左侧（LHS）之间的差值，结果见图 4，并对多个图库集取平均值。两者之间的强相关性（皮尔逊系数为 0.77）和较大的正斜率（4.75）证实，理论差值的增大与现实场景性能的提升相吻合，这表明尽管定理 1对分数分布作出了假设，其预测能力依然可靠。

1:1 评估。 尽管标准部署假设可以访问所有已注册受试者，但在某些实际场景中验证仅限于单个可用受试者。为评估 LocalScore 在这一限制下是否仍然有效，我们在 1:1 设置下对其进行评估，其中每次判定都涉及一个探针和一个图库主体。在此设置中，我们的方法仅使用当前受试者的图库样本来增强相似度分数（示意图见补充材料）。该配置的结果在表 2 中以我们的方法*报告。在不同模态和数据集上，即使无法访问完整图库，LocalScore 仍能取得出色的验证和开放集性能。令人意外的是，其开放集结果往往优于可访问所有受试者时的结果。例如，在 CCPG [36] 使用 Big-Gait [64], 时，LocalScore 达到 36.29%FNIR@0.1%FPIR，而不使用我们的方法时为 46.95%，使用整个图库时则为 37.04%。在验证方面，尽管性能通常低于使用完整图库时的设置，LocalScore 仍能显著优于基线评分。值得注意的是，在 Gait3D [71] 搭配 SwinGait [11], 时，其正确接受率@0.1%错误接受率从57.81% 提升至 65.62%；而使用整个图库时仅达到 63.65%。这些结果表明，LocalScore 能够有效地建模受试者内变化，即使只能访问单个受试者的图库，也能实现可靠验证。

人体识别。 我们在多模态BRIAR基准上评估了 LocalScore 在四种模态下的表现：人脸、步态、行人重识别（reID）和全身识别。每种模态均使用各自的特征提取器，结果汇总于表 3。LocalScore 持续提升开放集性能，在 CCVID 上的提升尤为显著。例如，它将行人重识别的 FNIR@1%FPIR 从 52.77 降至 40.49 ，并将步态的 TAR@0.1%FAR 从 50.72% 提升至 73.98%。在更具挑战性的 BRIAR [6]数据集上，LocalScore 也几乎在所有情况下都提升了性能。这些结果表明，LocalScore 能够无缝集成到现实场景系统中，不仅提升单个模型的表现，还能增强复杂的多模态人体识别流程。

与多样本方法的比较。我们将评分函数与标准的多样本聚合策略进行比较，这些策略会将多个图库样本合并为一个分数。具体而言，我们评估了三种方法：均值、最大值和最小值 [4]。如表4所示，这些方法呈现出一致的排序：我们的方法性能最强，其次是最大值，均值和最小值则相对较差。例如，在 CCPG [36], 上，我们的方法 FNIR 最低（37.04%@0.1%FPIR），相比之下，最大值为42.87%，均值为51.08%，最小值为83.23%。验证性能也呈现出类似趋势：我们的方法 TAR@0.1%FAR 达到92.95%，优于最大值（86.94%）、均值（79.22%）和最小值（48.65%）。总体而言，将局部几何信息纳入评分函数，能够在以下两种条件下实现更可靠的验证

模态	模型	数据集	图库	评分	FNIR@FPIR ↓		TAR@FAR ↑	
					1%	5%	0.1%	1%
Face	KPRPE [29]	IJB-S [25]	Avg	N/A	**94.11**	88.64	10.69	25.47
				我们的方法*	94.17	**88.00**	11.33	25.91
				Ours	94.49	89.12	**21.76**	**45.47**
			EVM [17]	N/A	**88.20**	**83.43**	18.93	31.83
				我们的方法*	94.35	83.94	17.76	31.28
				Ours	94.08	84.63	**22.93**	**36.83**
Gait	SwinGait [11]	Gait3D [71]	Avg	N/A	81.84±1.76	66.09±0.69	57.81	78.95
				我们的方法*	**78.59±2.04**	**61.76±0.94**	**65.62**	**86.47**
				Ours	78.80±1.99	62.98±0.78	63.65	79.08
			EVM [17]	N/A	83.63±1.54	62.40±1.26	69.68	88.61
				我们的方法*	**83.36±1.55**	**61.95±1.23**	69.89	**92.20**
				Ours	83.38±1.53	62.19±1.21	**79.84**	88.72
	BigGait [64]	CCPG† [36]	Avg	N/A	46.95±2.02	34.53±1.96	82.46	95.64
				我们的方法*	**36.39±2.77**	**24.10±2.12**	87.86	97.04
				Ours	37.04±2.71	25.34±1.99	**92.95**	**97.01**
			EVM [17]	N/A	46.64±3.93	34.94±3.41	86.25	97.37
				我们的方法*	**41.93±4.14**	**30.23±3.23**	87.13	97.10
				Ours	41.99±4.13	30.41±3.21	**93.29**	**97.78**
ReID	CAL [16]	CCVID [16]	Avg	N/A	39.98±0.88	34.38±0.91	68.19	81.81
				我们的方法*	**38.66±0.85**	**33.31±0.93**	69.24	82.97
				Ours	38.98±0.77	34.19±0.84	**75.23**	**85.34**
			EVM [17]	N/A	37.48±2.32	**28.72±1.05**	75.21	86.12
				我们的方法*	**37.30±2.06**	30.04±0.88	72.53	82.99
				Ours	37.48±1.99	30.98±0.82	**78.00**	**86.26**

表2. 真实世界生物特征数据集上的开放集 FNIR@FPIR 与验证 TAR@FAR。针对 FNIR@FPIR 报告了 95% 置信区间。我们的方法*表示 1:1 设置下的 LocalScore。†FNIR@FPIR 在 0.1% 和 1% 阈值下报告。

Task	模型	Ours	FNIR@FPIR ↓		TAR@FAR ↑	
			CCVID	BRIAR	CCVID	BRIAR
Face	KPRPE [29]	✗	12.99	93.6	91.12	14.7
		✓	**10.67**	**92.9**	**93.88**	**17.3**
Gait	BigGait [64]	✗	71.05	84.3	50.72	35.3
		✓	**59.92**	**84.0**	**73.98**	**37.7**
ReID	CLIP3D-行人重识别 [39]	✗	52.77	**86.8**	62.11	28.2
		✓	**40.49**	87.2	**80.46**	**32.2**
整体 Body	FarSight [38]	✗	15.41	76.3	87.41	39.2
		✓	**10.74**	**73.4**	**91.73**	**53.5**

表3. BRIAR 上的开放集 FNIR@FPIR、验证 TAR@FAR [6] 和 CCVID [16] 数据集。

Task	模型	数据集	方法	FNIR@FPIR ↓		TAR@FAR ↑	
				0.1%	1%	0.1%	1%
Gait	BigGait [64]	CCPG [36]	Max	42.87	28.92	86.94	**97.11**
			Min	83.23	70.97	48.65	77.88
			Mean	51.08	38.12	79.22	95.20
			R-R	47.84	36.56	81.12	95.14
			Ours	**37.04**	**25.34**	**92.95**	97.01
ReID	CAL [16]	CCVID† [16]	Max	40.55	**32.89**	69.65	82.83
			Min	54.49	45.86	60.97	80.36
			Mean	**38.74**	34.38	69.16	82.87
			R-R	40.52	34.12	68.88	83.60
			Ours	38.98	34.19	**75.23**	**85.34**

表 4. 多样本方法的开放集 FNIR@FPIR 和验证 TAR@FAR。†FNIR@FPIR 在 1% 和 5% 阈值下报告。 [R-R：重排序 [72]]

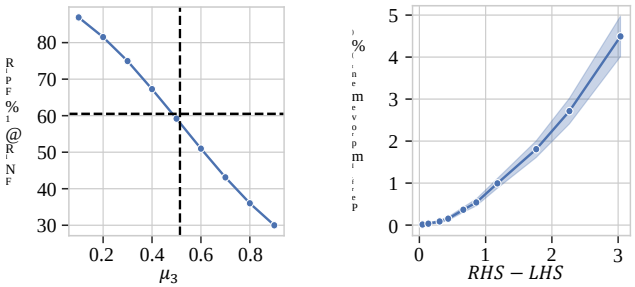
开放集和封闭集条件下，其表现优于忽略受试者内变化的基于池化的方法。

与重排序的比较。 为进行公平比较，我们将 LocalScore 与一种逐探针变体的重排序算法进行对比，该算法 [72],经过调整，可独立作用于每个探针。经典的重排序利用互惠邻居——即查询集合中的互为最近邻关系——因此通常需要多个探针才能发挥作用。如表 4所示，LocalScore 始终以显著优势超过逐探针重排序，未凸显了其适用于真实世界生物特征识别场景的特点，在这些场景中，探测样本通常单独到达。

5. 蒙特卡洛模拟

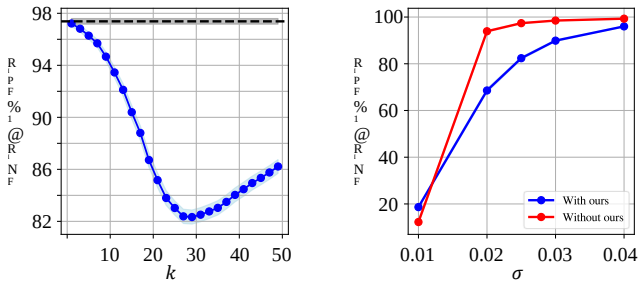
5.1 定理1

我们进行了蒙特卡洛模拟，以评估定理的实际有效性。1 我们在其理论假设和更一般的条件下生成随机



(a) (b)

图5。（a） μ_3 对开放集FNIR@1%FPIR的影响。垂直线表示理论预测的 μ_3^*，水平线表示未使用我们的方法时的性能。（b）使用我们的方法时，式（7）左右两侧的差值对FNIR@FPIR改善幅度的影响。阴影区域为99% 置信区间。



(a) k的影响。 (b) σ的影响。

图 6. k 和 σ 对开放集FNIR@1%FPIR的影响。(a)中，水平线表示未使用我们的方法时的性能。阴影区域为99%置信区间。

与定理假设一致的分数矩阵。1其中匹配、非匹配、配对KNN 和非同人分数均服从正态分布。随后，我们计算 μ_3^*，即使公式（7）两边相等的值。在固定其他所有分布参数的同时，我们改变 μ_3，并在图5a中观察到，当 $\mu_3 < \mu_3^*$ 时，性能会下降；而当 $\mu_3 > \mu_3^*$时，性能会提升。这种模拟与理论之间的一致性证实了定理1的预测有效性，即能够识别我们的方法何时会带来性能提升。

为考虑理论假设可能存在的偏差，我们将模拟扩展到随机生成的特征向量及其对应的分数矩阵。我们考察性能提升与式（7）右侧和左侧之间数值差距的关系；差距越大，匹配与非匹配k最近邻分数的区分度越高，因此性能提升也越显著。

我们创建包含匹配样本和非匹配样本的合成数据集，以评估我们的方法。具体而言，将 N 个类别中的20%指定为非匹配类别，其中仅包含探测样本。对于第 i个类别，数据取自一个多元高斯分布：

$$\mathcal{N}\left(\left[\cos\left(\tfrac{2\pi i}{N}\right), \sin\left(\tfrac{2\pi i}{N}\right)\right]^T, \sigma^2\mathbf{I}\right), \tag{9}$$

从而使类别均匀分布在单位圆上。匹配类别在图库和探测集中均包含 M 个样本，而非匹配类别仅包含 M 个探测样本。此处，σ 控制类内变化，N 表示类别总数。为确保稳健性，每项实验均使用不同的随机种子重复，000次。

图。5b 表明，LocalScore 能够在不同的 σ 和 k 条件下可靠地区分同源探针与非匹配探测样本。右侧始终高于左侧，这与大多数情况下观察到的开放集性能提升一致。此外，数值差异越大，对应的性能提升也越大，这表明尽管其假设数据服从正态分布，定理。 1 具有良好的泛化能力，并能有效预测我们的方法何时会提升 FNIR@FPIR。

5.2. 消融研究

我们使用与上一小节相同的设置，对超参数 k 和类内变化参数 σ 进行消融，并报告它们对FNIR@1% FPIR的影响。关于式（8）和TAR@FAR的补充分析见补充材料。

.....的影响 k。如图6a所示，当 k 较小时，开放集FNIR@1%FPIR会急剧下降，随后逐渐上升。这一趋势与现实观察相符，LocalScore始终优于基线。与模拟相比，实践中的最优 k 更小——$k = 1$ 在我们的实验中。虽然模拟中的高斯假设可能无法充分刻画现实世界的特征分布，但我们的方法始终能够提升开放集性能，在最优 k时取得超过15%的提升。

.....的影响 σ。如图6b所示，性能提升很大程度上取决于类内变化。对于较小的 σ （变化较低），提升幅度较小；但随着 σ 增大，LocalScore会迅速超越原始分数，在具有挑战性的情形下带来显著提升。这一模式与真实数据集相符：在步态和行人重识别等受试者内变化较大的模态中，收益更为明显。在变化较低的人脸识别任务中，LocalScore仍能带来显著提升，尤其是在验证方面。

6. 结论

我们提出一种简单而有效的算法，通过利用受试者内部的变化来提升开放集搜索和验证性能。我们在各种生物特征识别任务中评估了该方法，采用不同架构的模型，并使用多种损失函数进行训练，以展示其有效性。此外，我们还提供了实证和理论分析，以了解该方法在何种条件下能够提升性能。

致谢。 本研究部分由美国国家情报总监办公室（ODNI）下属的情报高级研究计划局（IARPA）通过 2022-21102100004 项目资助。本文所述观点和结论仅代表作者观点，不应被解释为必然代表 ODNI、IARPA 或美国政府明示或暗示的官方政策。尽管其中包含任何版权声明，美国政府仍有权为政府用途复制和分发转载件。

参考文献

[1] 白松、唐鹏、Philip H.S. Torr 和 Longin Jan Latecki。通过度量融合进行重排序，以实现目标检索和行人重识别。发表于计算机视觉与模式识别会议，2019年。3[2] Arko Barman 和 Shishir K Shah。Shape：一种用于行人重识别系统中基于共识的决策的新型图论算法。发表于国际计算机视觉大会，2017年。3[3] Abhijit Bendale 和 Terrance E Boult。迈向开放集深度网络。发表于计算机视觉与模式识别会议，2016年。2[4] Lacey Best-Rowden、韩虎、Charles Otto、Brendan F Klare 和 Anil K Jain。无约束人脸识别：从媒体集合中识别重点关注人员。《IEEE信息取证与安全汇刊》，2014年。6[5] 陈克奇、Vinkle Srivastav、Didier Mutter 和 Nicolas Padoy。从同步中学习：复杂场景下自监督的未校准多视角人员关联。发表于计算机视觉与模式识别会议，2025年。2[6] David Cornett、Joel Brogan、Nell Barber、Deniz Aykac、Seth Baird、Nicholas Burchfield、Carl Dukes、Andrew Duncan、Regina Ferrell、Jim Goddard 等。将准确的人员识别拓展至新的高度和距离：BRIAR数据集。发表于IEEE冬季计算机视觉应用大会，2023年。4, 5, 6, 7[7] Steve Cruz、Ryan Rabinowitz、Manuel Günther 和 Terrance E. Boult。实用开放集识别与后最大值精炼。发表于欧洲计算机视觉会议，2024年。2[8] 崔振宇、周家欢和彭宇新。差异化知识巩固：用于服装混合型终身行人重识别。发表于计算机视觉与模式识别会议，2025年。2[9] Saeed Ebrahimi、Sahar Rahimi、Ali Dabouei、Srinjoy Das、Jeremy M Dawson 和 Nasser M Nasrabadi。GIF：大规模人脸识别的生成式启发。发表于计算机视觉与模式识别会议论文集，第3528–3539页，2025年。2[10] Gustav Elfving。正态总体样本中极差的渐近分布。《生物统计学》，1947年。2[11] Chao Fan、侯赛辉、黄永祯和Shiqi Yu。探索用于实用步态识别的深度模型。arXiv预印本 arXiv:2303.03301，2023年。2, 5, 6, 7, 3, 4[12] Ronald Aylmer Fisher 和 Leonard Henry Caleb Tippett。样本中最大或最小值频率分布的极限形式。发表于《剑桥哲学学会数学汇刊》，1928年。1[13] 耿传兴、黄盛军和陈松灿。开放集识别的最新进展：综述。IEEE《模式分析与机器智能汇刊》，2020年。2[14] Boris Gnedenko. 关于随机级数最大项的极限分布。《数学年刊》，1943年。1[15] Patrick Grother、Mei Ngan和Kayee Hanaoka。人脸识别供应商测试（FRVT）第2部分：身份识别，2019年。2[16] Xinqian Gu、Hong Chang、Bingpeng Ma、Shutao Bai、Shiguang Shan和Xilin Chen。仅使用RGB模态的换装行人重识别。载于计算机视觉与模式识别会议，2022年。2、5、7、3、4[17] Manuel G¨unther、Steve Cruz、Ethan M Rudd和Terrance E Boult。面向开放集人脸识别。载于计算机视觉与模式识别会议研讨会，2017年。2、3、6、7[18] Manuel G¨unther、Peiyun Hu、Christian Herrmann、Chi-Ho Chan、Min Jiang、Shufan Yang、Akshay Raj Dhamija、Deva Ramanan、J¨urgen Beyerer、Josef Kittler 等。无约束人脸检测与开放集人脸识别挑战赛。载于国际生物识别联合会议。IEEE，2017年。1[19] Gavriel Habib、Noa Barzilay、Or Shimshi、Rami Ben-Ari和Nir Darshan。Cargait：面向步态识别的基于交叉注意力的重排序。载于国际计算机视觉大会，2025年。2[20] Alexander Hermans、Lucas Beyer和Bastian Leibe。为行人重识别中的三元组损失辩护。arXiv预印本 arXiv:1703.07737，2017年。2[21] Panjian Huang、侯赛辉、Junzhou Huang和黄永祯。学习用于步态识别的统一模板。载于国际计算机视觉大会，2025年。2[22] Anil K Jain、Debayan Deb和Joshua J Engelsma。生物特征识别：信任，但需验证。《IEEE生物特征、行为与身份科学汇刊》，2021年。2[23] Anil K. Jain、Arun A. Ross、Karthik Nandakumar和Thomas Swearingen。人脸识别，第119–173页。施普林格国际出版公司，2025年。1[24] Dongyang Jin、Chao Fan、Jingzhe Ma、Jingkai Zhou、Weihua Chen和Shiqi Yu。论用于步态识别的行走视频去噪。载于计算机视觉与模式识别会议，2025年。2[25] Nathan D Kalka、Brianna Maze、James A Duncan、Kevin O’Connor、Stephen Elliott、Kaleb Hebert、Julia Bryan和Anil K Jain。IJB-S：情报高级研究计划局Janus监控视频基准。载于生物识别：理论、应用与系统研讨会，2018年。5、7、3、4[26] Wajahat Khalid、Bin Liu、Xulin Li、Muhammad Waqas和Muhammad Sher Afgan。连接天空与地面：面向空地行人重识别的视角不变特征学习。载于国际计算机视觉大会，2025年。2[27] Minchul Kim、Anil K Jain和Xiaoming Liu。AdaFace：面向人脸识别的质量自适应边界。载于计算机视觉与模式识别会议，2022年。2[28] Minchul Kim、Feng Liu、Anil K Jain和Xiaoming Liu。聚类与聚合：面向大规模探针集的人脸识别。载于神经信息处理系统大会，2022年。2、3、6、5[29] Minchul Kim、Yiyang Su、Feng Liu、Anil Jain和Xiaoming Liu。面向人脸识别的关键点相对位置编码。载于计算机视觉与模式识别会议，2024年。4、7、3、5[30] Minsoo Kim、Min-Cheol Sagong、Gi Pyo Nam、Junghyun Cho和Ig-Jae Kim。VigFace：面向无隐私风险人脸识别数据集的虚拟身份生成。载于IEEE/CVF国际计算机视觉大会论文集，第10043–10053页，2025年。2

[31] Minchul Kim、Dingqiang Ye、Yiyang Su、Feng Liu和Xiaoming Liu。Sapiensid：人体识别的基础。载于计算机视觉与模式识别会议，2025年。2[32] Sunpill Kim、Seunghun Paik、Chanwoo Hwang、Dongsoo Kim、Junbum Shin和Jae Hong Seo。Idface：面向高效安全身份识别的人脸模板保护。载于计算机视觉与模式识别会议，2025年。2[33] Shu Kong和Deva Ramanan。Opengan：通过开放数据生成实现开放集识别。载于国际计算机视觉大会，2021年。2[34] Hari Chandana Kuchibhotla、Sumitra S Malagi、Shivam Chandhok和Vineeth N Balasubramanian。稍后出现未见类别？没问题。载于计算机视觉与模式识别会议，2022年。2[35] Longhua Li、Lei Qi和Xin Geng。面向可扩展行人重识别的单样本知识迁移。载于国际计算机视觉大会，2025年。2[36] Weijia Li、侯赛辉、Chunjie Zhang、曹春水、刘旭、黄永祯和Yao Zhao。换装条件下行人重识别与步态识别的深入探索。载于计算机视觉与模式识别会议，2023年。1、5、6、7、3、4[37] Xin Liang和Yogesh S Rawat。Differ：通过语义线索解耦身份特征以用于换装行人再识别。载于计算机视觉与模式识别会议，2025年。2[38] Feng Liu、Ryan Ashbaugh、Nicholas Chimitt、Najmul Hassan、Ali Hassani、Ajay Jaiswal、Minchul Kim、Zhiyuan Mao、Christopher Perry、Zhiyuan Ren等。FarSight：远距离与高海拔条件下物理驱动的全身生物特征识别系统。载于IEEE冬季计算机视觉应用大会，2024年。1、2、7[39] Feng Liu、Minchul Kim、Zhiyuan Ren和Xiaoming Liu。利用双重引导蒸馏CLIP以学习判别性人体形状表征。载于计算机视觉与模式识别会议，2024年。2、7[40] Feng Liu、Nicholas Chimitt、Lanqing Guo、Jitesh Jain、Aditya Kane、Minchul Kim、Wes Robbins、Yiyang Su、Dingqiang Ye、Xingguang Zhang等。高度与距离条件下的人员识别：人脸、人体形状与步态的融合。arXiv预印本arXiv:2505.04616，2025年。2[41] Jianbo Ouyang、Hui Wu、Min Wang、Wengang Zhou和Houqiang Li。面向视觉重排序的基于自注意力的上下文相似度聚合。载于神经信息处理系统大会，2021年。3[42] Zhiqi Pang、Junjie Wang、Lingling Zhao和Chunyu Wang。面向无监督换装行人重识别的身份-服装相似度建模。载于计算机视觉与模式识别会议，2025年。2[43] Zhengyuan Peng、Jianqing Xu、Yuge Huang、Jinkun Hao、Shouhong Ding、Zhizhong Zhang、Xin Tan和Lizhuang Ma。Stylized-face：面向人脸识别的百万级风格化人脸数据集。载于国际计算机视觉大会，2025年。2[44] Norman Poh、Samy Bengio和Jerzy Korczak。面向生物特征认证的多样本多源模型。载于第12届IEEE信号处理神经网络研讨会论文集。IEEE，2002。2[45] Zhiyuan Ren、Yiyang Su和Xiaoming Liu。由ChatGPT驱动的面向图像分类的层次化比较。神经信息处理系统进展，36：69706–69718，2023年。2[46] Arun Ross、Sudipta Banerjee、Cunjian Chen、Anurag Chowdhury、Vahid Mirjalili、Renu Sharma、Thomas Swearingen和Shivangi Yadav。生物特征识别中的若干研究问题：未来在召唤。载于ICB，2019年。1[47] Arun A Ross、Anil K Jain 和 Karthik Nandakumar 《多生物特征识别手册》。施普林格，2006年。1[48] M Saquib Sarfraz、Arne Schumann、Andreas Eberle 和 Rainer Stiefelhagen。采用扩展跨邻域重排序的姿态敏感嵌入，用于行人重识别。载于计算机视觉与模式识别会议，2018年。3[49] 沈传福、Shiqi Yu、王继龙、George Q Huang 和王亮。深度步态识别综合综述：算法、数据集与挑战。arXiv 预印本 arXiv:2206.13732，2022年。1[50] 沈传福、王瑞、段立新和 Shiqi Yu。Lidar-gait++：从激光雷达点云中学习局部特征与尺寸感知，用于三维步态识别。载于计算机视觉与模式识别会议，2025年。2[51] 沈曦、肖扬、Shell Xu Hu、Othman Sbai 和 Mathieu Aubry。图像检索与传导式少样本分类的重排序。载于神经信息处理系统大会，2021年。3[52] Yiyang Su、Minchul Kim、Feng Liu、Anil Jain 和 Xiaoming Liu。开放集生物特征识别：超越优秀的闭集模型。载于欧洲计算机视觉会议，2024年。2，3，5[53] Yiyang Su、施云平、Feng Liu 和 Xiaoming Liu。Hamobe：面向基于视频的行人再识别的分层自适应生物特征专家混合。载于国际计算机视觉大会，2025 2。孙[54]一游、明一飞、朱小进和李一轩。利用深度近邻进行分布外检测。载于国际机器学习会议，2022年。3[55] 谭福文、袁江波和 Vicente Ordonez。使用重排序变换器进行实例级图像检索。载于国际计算机视觉大会，2021年。3[56] Sagar Vaze、韩凯、Andrea Vedaldi 和 Andrew Zisserman。开放集识别：优秀的闭集分类器就已足够。载于国际学习表征会议，2022年。2[57] 王世宁、Yunlong Wang、吴睿琪、焦炳良、王文轩和王鹏。Secap：面向空地网络中跨视角行人重识别的自校准自适应提示。载于计算机视觉与模式识别会议，2025年。2[58] 王子泰、Qi Qian、杨志勇、何源、Xiaochun Cao 和黄庆明。OpenAUC：迈向面向AUC的开放集识别。载于神经信息处理系统大会，2022年。2[59] 王增斌、侯赛辉、李俊杰、刘旭、曹春水、黄永祯、王思野和张曼。Gait-x：探索用于广义步态识别的 X 模态。载于国际计算机视觉大会，2025年。2[60] 徐昆仑、卓帆、李江萌、邹旭和Jiah uan Zhou。通过双知识协作进行自强化原型演化，用于半监督终身行人再识别。载于国际计算机视觉大会，2025年。2[61] 杨景康、王鹏云、邹德健、周子棠、丁坤元、彭文轩、王浩琦、陈光耀、李博、孙一游等。Openood：广义分布外检测基准。载于神经信息处理系统大会，2022年。3[62] 杨锦熙、李赫、杜博和叶茫。Cheb-gr：重新思考行人重识别重排序中的 K近邻搜索。载于计算机视觉与模式识别会议，2025年。3[63] 杨少鹏、王继龙、侯赛辉、刘旭、曹春水、王亮和黄永祯。连接步态识别与大语言模型序列建模。载于计算机视觉与模式识别会议，2025年。2

[64] Dingqiang Ye、Chao Fan、Jingzhe Ma、Xiaoming Liu 和 Shiqi Yu。BigGait：利用大型视觉模型学习所需的步态表征。载于计算机视觉与模式识别会议，2024年。2，5，6，7，3，4[65] Dingqiang Ye、Chao Fan、Zhanbo Huang、Chengwen Luo、Jianqiang Li、Shiqi Yu 和 Xiaoming Liu。Biggergait：利用大型视觉模型的逐层表征，释放步态识别潜能。载于神经信息处理系统大会，2025年。2[66] Jinghan You、Shanglin Li、Yuanrui Sun、Jiangchuan Wei、Mingyu Guo、Chao Feng 和 Jiao Ran。Lvface：面向人脸识别的大型视觉模型渐进式聚类优化。载于国际计算机视觉大会，2025年。2[67] Chao Yuan、Guiwei Zhang、Changxiao Ma、Tianyi Zhang 和 Guanglin Niu。从姿态到身份：通过特征中心化实现无需训练的行人重识别。载于计算机视觉与模式识别会议，2025年。2[68] Asmat Zahra、Nazia Perwaiz、Muhammad Shahzad 和 Muhammad Moazam Fraz。行人重识别：针对特定领域的开放挑战与未来趋势回顾。《模式识别》，2023年。1[69] Xiao-Wen Zhang、Delong Zhang、Yi-Xing Peng、Zhi Ouyang、Jingke Meng 和 Wei-Shi Zheng。Viperson：灵活生成用于行人重识别的虚拟身份。载于国际计算机视觉大会，2025年。2[70] Yuqi Zhang、Qi Qian、Hongsong Wang、Chong Liu、Weihua Chen 和 Fan Wang。基于图卷积的视觉检索高效重排序。《IEEE多媒体汇刊》，2023年。3[71] Jinkai Zheng、Xinchen Liu、Wu Liu、Lingxiao He、Chenggang Yan 和 Tao Mei。利用稠密三维表征和基准实现野外步态识别。载于计算机视觉与模式识别会议，2022年。5，6，7，3，4[72] Zhun Zhong、Liang Zheng、Donglin Cao 和 Shaozi Li。利用 k-倒数编码进行行人重识别重排序。载于计算机视觉与模式识别会议，2017年。3，7，4[73] Jie Zhu、Yiyang Su、Minchul Kim、Anil Jain 和 Xiaoming Liu。面向人体识别的质量引导分数融合专家混合框架。载于国际计算机视觉大会，2025年。2

局部评分：用于生物特征识别的局部密度感知相似度评分

补充材料

A. 对Rank-1准确率的影响

算法 1 对 Rank-1 准确率的影响可由以下定理概括：

定理2. 设 s_1、s_2、……、s_n 表示由某个闭集识别算法生成的一组相似度分数，其中 s_i 表示探针与第 i个图库主体之间的相似度分数。若算法*1* 将一个非负值d 加到最高分上，则任意排名下的闭集准确率都不会改变。

证明。 为此，我们不失一般性地假设分数按非递减顺序排列，即，$s_1 \leq s_2 \leq \cdots \leq s_n$。根据算法 1，最高分 s_n被调整为 $s_n + d$，其中 $d \geq 0$。由此得到以下不等式：

$$s_1 \leq s_2 \leq \cdots \leq s_n \leq s_n + d. \tag{10}$$

此调整保持了分数的相对排名。

排名 r 处的闭集准确率由得分排名前 r 的正确匹配数量决定。假设正确匹配对应的分数为 s_j。若 s_j 最初位于前 r 名之内，则有 $j \geq n + 1 - r$。由于此调整保持相对排名，s_j 仍正确地位于前 r 名之内。反之，若 s_j 最初不在前 r 名得分之中，则有 $j < n+1-r$。s_n 的增加不会影响低于 s_n 的其他分数的排名，从而确保s_j 仍不会排入前 r 名。

这两种情况下，我们都可以得出结论：算法 1 不会改变任何排名处的闭集准确率。证毕。

B. 定理1的证明

本节将证明定理1。证明之前，我们先重述该定理，供您参考。

假设每个对象的真实分数服从正态分布 $\mathcal{N}(\mu_1, \sigma_1)$。每个对象的冒名者分数服从分布 $\mathcal{N}(\mu_2, \sigma_2)$。同源探针的 KNN 分数服从分布 $\mathcal{N}(\mu_3, \sigma_3)$，而非同源探针的 KNN 分数服从分布 $\mathcal{N}(\mu_4, \sigma_4)$。此外，设 N_1为同源探针的数量，N_2 为非同源探针的数量，M 为图库受试者的数量，N_3 为冒名者分数的总数，r_1 和 r_2 为可容忍的每探针最大非同源分数数量（例如，0.1% $\times N_2$）和非匹配分数数量（例如，0.1% $\times N_3$）。

定理1。 算法 *1*在期望阈值下改善了开放集 *FNIR@FPIR*

$$\frac{\mu_2 + \mu_4 + \delta\sqrt{\sigma_2^2 + \sigma_4^2} - (\mu_1 + \mu_3)}{\sqrt{\sigma_1^2 + \sigma_3^2}} < \frac{(\mu_2 + \delta\sigma_2) - \mu_1}{\sigma_1}, \tag{5}$$

其中 $\delta = -\ln(-\ln(1-r_1/N_2))$，并且在满足以下条件时，可提高期望阈值下的验证 *TAR@FAR*：

$$\frac{\mu_2 + \mu_4 + \Phi^{-1}\left(\frac{N_3 - r_2 - \alpha}{N_3 - 2\alpha + 1}\right)\sqrt{\sigma_2^2 + \sigma_4^2} - (\mu_1 + \mu_3)}{\sqrt{\sigma_1^2 + \sigma_3^2}} < \frac{\mu_2 + \Phi^{-1}\left(\frac{N_3 - r_2 - \alpha}{N_3 - 2\alpha + 1}\right)\sigma_2 - \mu_1}{\sigma_1}, \tag{6}$$

其中 $\Phi^{-1}(\cdot)$ 是正态分布累积分布函数的反函数，$\alpha = \pi/8$ 是一个常数。

证明。 我们分别证明定理的两部分，即式和式 (8)，并从 (7) 式 (7) 开始。

为证明式 (7)，我们先考虑不含 KNN 分数时的 FNIR 阈值。根据 Fisher-Tippett-Gnedenko 定理 [12, 14], ，由于非匹配分数服从高斯分布，最大非同人分数会收敛到耿贝尔分布。

已知不含 KNN 分数时非匹配分数的均值μ_2 及其标准差为 σ_2，相应的阈值由耿贝尔分布的分位数确定。具体而言，FPIR 等于 $\frac{r_1}{N_2}$ 时的阈值为：

$$\mu_2 + \left(-\ln - \ln\left(1 - \frac{r_1}{N_2}\right)\right) = \mu_2 + \delta\sigma_2, \tag{11}$$

若令 $\delta = -\ln\left(-\ln\left(\frac{r_1}{N_2}\right)\right)$，则该值是对应于所需 FPIR 的临界值。此阈值由耿贝尔分布的逆累积分布函数（CDF）得出。

接下来，为了计算此阈值下的 FNIR，我们利用匹配分数服从均值为 μ_1、标准差为 σ_1 的高斯分布这一事实。FNIR 可由阈值处匹配分数的互补累积分布函数给出：

$$\text{FNIR} = 1 - \Phi\left(\frac{(\mu_2 + \delta\sigma_2) - \mu_1}{\sigma_1}\right), \tag{12}$$

其中，Φ 是标准正态分布的累积分布函数。这是基于原始匹配分数，在该阈值下的预期 FNIR。

现在，我们引入 KNN 分数。使用 KNN 分数时，非匹配分数变为两个高斯分布之和：一个均值为 μ_2、方差为 σ_2^2，另一个均值为 μ_4、方差为 σ_4^2。得到的分布是高斯分布，其均值为 $\mu_2+\mu_4$，标准差为 $\sqrt{\sigma_2^2+\sigma_4^2}$。

因此，使用 KNN 分数的 FPIR 阈值由下式给出

$$\mu_2+\mu_4+\delta\sqrt{\sigma_2^2+\sigma_4^2}. \tag{13}$$

类似地，在预期阈值处使用 KNN 分数的 FNIR 为

$$\text{FNIR}_{\text{KNN}} = 1-\Phi\left(\frac{\mu_2+\mu_4+\delta\sqrt{\sigma_2^2+\sigma_4^2}-(\mu_1+\mu_3)}{\sqrt{\sigma_1^2+\sigma_3^2}}\right), \tag{14}$$

其中 $\mu_1+\mu_3$ 是使用 KNN 的匹配分数的均值，而 $\sqrt{\sigma_1^2+\sigma_3^2}$ 是标准差。为简便起见，此处假设将 KNN 分数加到所有匹配分数上，这对于现实世界的生物识别系统而言是一个合理的假设。

为证明该定理的第一部分，我们证明算法 1改进了 FNIR。这等价于证明使用 KNN 分数的 FNIR 小于不使用 KNN 分数的 FNIR。为此，我们比较两个 FNIR 表达式中累积分布函数内部的自变量。具体而言，我们要求：

$$\frac{\mu_2+\mu_4+\delta\sqrt{\sigma_2^2+\sigma_4^2}-(\mu_1+\mu_3)}{\sqrt{\sigma_1^2+\sigma_3^2}} < \frac{(\mu_2+\delta\sigma_2)-\mu_1}{\sigma_1}. \tag{15}$$

该不等式保证，与不使用 KNN 分数的 FNIR 相比，使用 KNN 分数的 FNIR 更小（得到改进）。由于 $\Phi(\cdot)$ 是单调递增函数，该不等式意味着使用 KNN 分数的 FNIR 确实更低。

接下来，我们讨论定理的第二部分：在期望阈值处提升验证 TAR@FAR。我们首先推导非匹配分数的 FAR 阈值，这些分数服从均值为 μ_2、标准差为 σ_2 的高斯分布。根据 [10], ，给定错误接受率 $\frac{r_2}{N_3}$ 时，FAR 阈值为：

$$\mu_2+\Phi^{-1}\left(\frac{N_3-r_2-\alpha}{N_3-2\alpha+1}\right)\sigma_2. \tag{16}$$

其中，$\alpha=\frac{\pi}{8}$ 是一个常数，Φ^{-1} 是标准正态累积分布函数的反函数。

对于验证的正确接受率，我们假设算法 1 将 KNN 分数加到所有高于此阈值的非匹配分数上。因此，新的预期验证阈值为：

$$\mu_2+\mu_4+\Phi^{-1}\left(\frac{N_3-r_2-\alpha}{N_3-2\alpha+1}\right)\sqrt{\sigma_2^2+\sigma_4^2}. \tag{17}$$

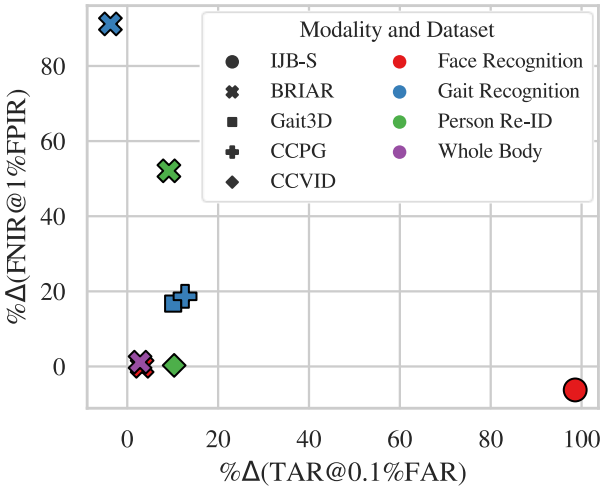


图 7. LocalScore 在不同模态和数据集上的相对性能提升（%Δ）计算方式为（我们的方法 − 基线）/基线，其中“我们的方法”表示使用 LocalScore 时的 TAR@0.1%FAR 或 FNIR@1%FPIR，“基线”指未使用 LocalScore 时的对应值。在（0，0）处，表示两个指标均未优于基线。这些结果源自表2，凸显了我们的方法带来的性能提升。

最后，我们比较加入和未加入 KNN 分数时验证阈值的 z 分数。匹配分数分布（未加入 KNN）的 z 分数为：

$$\frac{\mu_2+\mu_4+\Phi^{-1}\left(\frac{N_3-r_2-\alpha}{N_3-2\alpha+1}\right)\sqrt{\sigma_2^2+\sigma_4^2}-(\mu_1+\mu_3)}{\sqrt{\sigma_1^2+\sigma_3^2}}. \tag{18}$$

对于 KNN 分数的情况，z 分数为：

$$\frac{\mu_2+\Phi^{-1}\left(\frac{N_3-r_2-\alpha}{N_3-2\alpha+1}\right)\sigma_2-\mu_1}{\sigma_1}, \tag{19}$$

通过比较两个分布的 z 分数，我们得出结论：如果满足公式（1，算法可在期望阈值处提升验证 TAR@FAR8）。Q.E.D.

C. 补充实验

相对性能提升 图 7 以直观方式展示了表 2的结果：与不同模态和数据集上的基线方法相比，性能提升的百分比。LocalScore 在所有数据集上都实现了显著的性能提升，增强了验证任务中的 TAR@0.1%FAR 或 FNIR@1%FPIR。值得注意的是，它在人脸识别这一颇具挑战的 IJB-S 数据集上将验证性能提升了约 100%，并使 BRIAR 上的步态识别性能提升了约 90%。总体而言，LocalScore

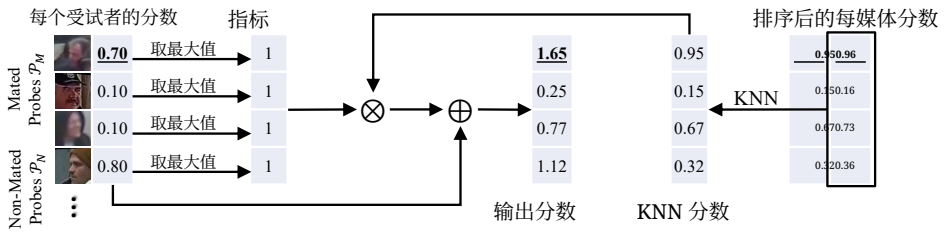


图8. 使用 LocalScore 进行单主体评估的示意图。与图 2相比，当图库中只有一个受试者时，指标始终为 1，因为每个探针只有一个主体得分。因此，将同一受试者的第 k大分数加到主体得分中。

模型	数据集	Mod.	C_i	% 画廊	Run-time	FNIR@FPIR ↓		TAR@FAR ↑	
						0.1%	1%	0.1%	1%
KPRPE [29]	IJB-S† [25]	Face	N/A	< 1	148	94.11	88.64	10.69	25.47
			5	6.3	285	**93.77**	88.81	21.88	45.27
			20	23.4	595	94.50	**88.57**	**22.14**	**45.62**
			∞	100		94.49	89.12	21.76	45.46
SwinGait [11]	Gait3D† [71]	Gait	N/A	< 1	124	81.84	66.09	57.81	78.95
			5	53.7	246	83.50	69.10	**63.65**	78.95
			20	92.3	319	81.01	65.43	**63.65**	**79.08**
			∞	100	338	**78.80**	**62.98**	**63.65**	**79.08**
BigGait [64]	CCPG [36]	Gait	N/A	< 1	29	46.95	34.53	82.46	95.64
			5	14.2	60	41.04	29.52	92.25	97.00
			20	56.8	175	39.23	27.34	92.89	97.00
			∞	100	301	**37.04**	**25.34**	**92.95**	**97.01**
CAL	CCVID† [16]	ReID	N/A	< 1	15	39.98	34.38	68.19	81.81
			5	60.8	31	**38.39**	**33.72**	**75.69**	**85.34**
			20	100	41	38.98	34.19	75.23	**85.34**
			∞	100	41	38.98	34.19	75.23	**85.34**

表5. 不同 C_i下的开放集 FNIR@FPIR 和验证 TAR@FAR。% 画廊：簇化图库大小与完整图库大小之比。运行时间以毫秒（ms）为单位报告。†FNIR@FPIR 在1%和5%的阈值下报告。

在步态识别和行人重识别方面展现出更大的性能增益，因为开放集生物特征识别在这些任务中面临更大的挑战，同时也提升了人脸识别性能。这些结果凸显了我们的方法在多种生物特征识别任务中的有效性和泛化能力。

C.1. 1:1 评估。

1:1 设置与标准评估设置之间的关键区别，在于对图库数据可用性的假设不同。标准评估遵循传统的生物特征识别协议，其中所有已注册主体的图库均可用于验证和检索。这反映了典型的运行场景：在部署之前，完整图库已预先注册；即使在验证过程中，非声称身份对象的图库也可用，尽管大多数系统并不会明确使用这些图库。然而，在资源极其受限的环境中，对无关对象图库的访问可能会受到限制。为了在此类情境下评估 LocalScore，我们还评估了 1:1 设置中的性能，在该设置下仅提供声称身份的图库样本。

况下，我们还评估了 1:1 设置中的性能，在该设置下仅提供声称身份的图库样本。

如图 8所示，该场景生成一个每个受试者分数列，反映唯一可访问的图库；与图2相比，由于排除未声称身份的媒体，因此每个媒体的得分更少。因此，所有指标均为1，因为每个指标都是通过对单个值取最大值计算得出的，并且 k-NN 分数会被加到每个受试者的分数中。尽管存在这些限制，表2显示，LocalScore 仍能提升多个生物识别模型在不同模态下的开放集检索和验证性能。

C.2. 图库聚类

表 5 表明，增加 C_i——这相当于使用更大比例的聚类图库——通常能提升所有模型在开放集和验证任务中的性能

Task	模型	数据集	Mod.	方法	FNIR@FPIR ↓		TAR@FAR ↑	
					0.1%	1%	0.1%	1%
Face	KPRPE [29]	IJB-S† [25]	Face	Max	95.39	89.80	8.73	22.40
				Min	99.83	99.11	1.40	7.45
				Mean	96.14	90.92	9.55	25.06
				R-R	95.32	92.45	8.69	23.98
				Ours	**94.49**	**89.12**	**21.76**	**45.47**
Gait	SwinGait [11]	Gait3D [71]	Gait	Max	81.92	**60.40**	70.66	91.11
				Min	88.02	73.24	50.99	78.34
				Mean	83.72	60.81	**74.00**	**93.73**
				R-R	82.84	67.61	56.96	77.98
				Ours	**78.80**	62.98	63.65	79.08
Gait	BigGait [64]	CCPG [36]	Gait	Max	42.87	28.92	86.94	**97.11**
				Min	83.23	70.97	48.65	77.88
				Mean	51.08	38.12	79.22	95.20
				R-R	47.84	36.56	81.12	95.14
				Ours	**37.04**	**25.34**	**92.95**	97.01
ReID	CAL [16]	CCVID† [16]	ReID	Max	40.55	**32.89**	69.65	82.83
				Min	54.49	45.86	60.97	80.36
				Mean	**38.74**	34.38	69.16	82.87
				R-R	40.52	34.12	68.88	83.60
				Ours	38.98	34.19	**75.23**	**85.34**

表 6. 多样本方法的开放集 FNIR@FPIR 和验证 TAR@FAR。†FNIR@FPIR 在 1% 和 5% 阈值下报告。[R-R：重排序 [72]]

以及模态，同时只带来适度的运行时间开销。在基线设置（C_i = N/A）中，使用的图库不足 1%，性能始终较低。随着 C_i 增加，FNIR 降低，正确接受率（TAR）提高；大多数模型在利用完整图库（$C_i = \infty$）时取得最佳或接近最佳的结果。这一趋势在步态和行人重识别模型中尤为明显：例如，CCPG 上的 BigGait，其 FNIR@0.1% 从 46.95% 降至 37.04%，而 TAR@0.1% 则从 82.46% 升至 92.95%。SwinGait 在 Gait3D 上以及 CAL 在 CCVID 上也有类似提升。对于 IJB-S 上的人脸识别，提升较早趋于平稳，$C_i = 20$ 已达到或超过完整图库设置的表现。总体而言，结果表明，扩大聚类图库能够持续提升不同生物特征模态的识别性能，同时额外的计算开销仍处于实际部署可承受的范围内。

C.3. 与多样本方法的比较

表 6 比较了 LocalScore 与传统多样本聚合策略在人脸、步态和行人重识别任务中的表现。总体而言，结果表明，最大值、均值和最小值等基于池化的方法表现不一致：最大值通常能取得有竞争力的 FNIR，但正确接受率相对较低；均值通常能提升验证性能，但在开放集条件下表现不佳。最小值在所有设置下表现最差，证实其对异常值较为敏感。重排序（R-R）带来了一定程度的提升，但仍受限于其对互惠近邻和多个探测样本的依赖。相比之下，LocalScore 带来了显著且稳定的提升，尤其是在具有挑战性的开放集场景中。例如，在 CCPG 步态任务中，LocalScore 将 FNIR@0.1% FPIR 降至 37.04%，显著优于最大值（42.87%）和均值（51.08%），同时还取得了最高的 TAR@0.1% FAR（92.95%）。在人脸和行人重识别设置中也能看到类似的提升，LocalScore 在所有指标上均取得最佳或接近最佳的表现。这些结果凸显了将局部图库结构纳入评分函数的优势，使其能够比传统聚合或重排序方法实现更可靠的识别。

C.4. 统计显著性检验

我们进行双样本t检验，以评估在随机匹配与非匹配划分中，LocalScore所带来的性能提升是否具有统计显著性，如表7所示。结果表明，步态识别中的所有改进均具有高度显著性，其中大多数 p值小于0.0001。尽管CAL模型在CCVID数据集上的改进显著性较低，LocalScore相较于CAL仍实现了显著的验证性能提升，证明了其有效性。在其他数据集上，大规模探针集大小降低了随机波动。

模型	数据集	图库	FNIR@FPIR		TAR@FAR	
			1%	5%	0.1%	1%
SwinGait	Gait3D	平均值+欧氏	0.0001	0.0001	0.0001	0.0001
		EVM	0.6441	0.3295	0.0001	0.3120
BigGait	CCPG	平均值+欧氏	0.0001	0.0001	0.0001	0.0001
		EVM	0.0167	0.0014	0.0001	0.0001
CAL	CCVID	平均值+余弦	0.0001	0.4077	0.0001	0.0001
		EVM	1.0000	0.0012	0.0001	0.6741

Table 7. FNIR@FPIR、TAR@FAR 及其在真实世界生物特征数据集上的统计显著性检验。

模态	模型	数据集	图库	后处理	FNIR@FPIR ↓		TAR@FAR ↑		R@1 ↑
					0.1%	1%	0.1%	1%	
Face	KPRPE [29]	BRIAR‡ [6]	CAFace [28]	N/A	65.55	54.93	59.93	**72.52**	84.34
				+常数。	65.55	55.00	61.64	72.39	84.34
				×2	65.55	55.00	61.64	**72.52**	84.34
				1,2-最近邻	64.99	**54.85**	61.82	**72.52**	84.34
				1,2,3-NN	65.12	54.91	**62.01**	**72.52**	84.34
				Ours	**64.81**	54.91	61.89	**72.52**	84.34
Gait	SwinGait [11]	Gait3D† [71]	平均值 + 欧几里得	N/A	81.84±1.76	66.09±0.69	57.81	78.95	72.2
				+常数。	81.85±1.75	66.46±0.64	**63.65**	**79.08**	72.2
				×2	81.85±1.75	66.46±0.64	**63.65**	**79.08**	72.2
				1,2-NN	80.19±1.85	64.69±0.61	**63.65**	**79.08**	72.2
				1,2,3-NN	81.78±1.65	66.34±0.61	**63.65**	**79.08**	72.2
				Ours	**78.80±1.99**	**62.98±0.78**	**63.65**	**79.08**	72.2

表8. 真实世界生物特征数据集上的开放集 FNIR@FPIR、验证 TAR@FAR 和封闭集 Rank-1 准确率。FNIR@FPIR 报告 95% 置信区间。†报告 FNIR@1%FPIR 和 FNIR@5%FPIR。‡根据官方评估协议报告 Rank-20 准确率 [键：R@1=Rank-1 准确率]

C.5. KNN 分数的影响

我们通过多种变体开展补充实验，以分析KNN分数对原始分数矩阵的影响。首先，我们为每个探针的最高分加上常数1，这不会在匹配探测样本与非匹配探测样本之间引入额外的可区分性（“+const.”）。其次，我们将每个探针的最高分替换为其两倍，在保持匹配探测样本与非匹配探测样本之间有限可区分性的同时增大其幅度，但不利用KNN分数（“×2”）。第三，我们使用2个最近邻（1,2-KNN）或3个最近邻（1,2,3-KNN）的平均值。

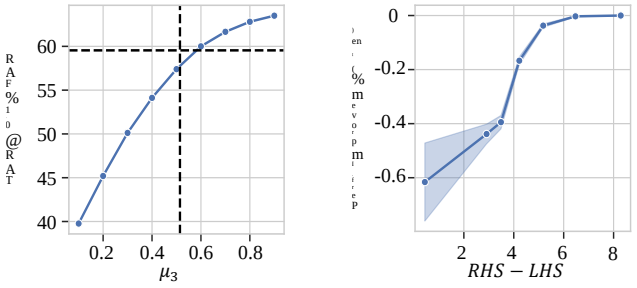
如表 8所示，尽管这些方法均提升了验证性能，但无一超越我们提出的方法。对于开放集 FNIR@FPIR，这些变体在 Gait3D 上均未优于 LocalScore。此外，在人脸识别中，“+const.”和“×2”方法未能提升开放集性能，这凸显了引入 KNN 信息以获得更好开放集结果的必要性。当所有近邻都能提供合理性能时，纳入更多最近邻会得到相近结果，正如人脸识别。然而，该策略在步态识别中表现更差，因为额外的最近邻贡献的是质量较低的信息。总体而言，我们得出结论：KNN 分数对于提升验证性能和开放集性能既至关重要，又行之有效。

D. 附加分析

D.1. 关于定理的附加分析1

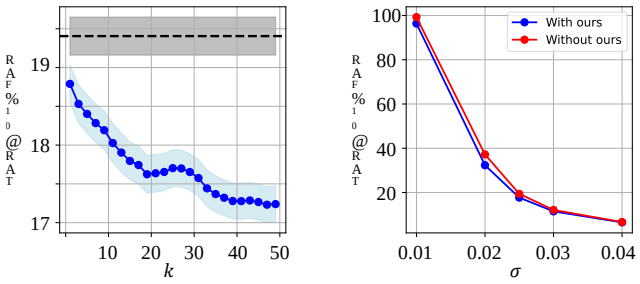
基于图5中的分析，我们考察公式（8）如何为现实世界验证性能提供洞见，重点关注 μ_3的影响。使用与图5a相同的实验设置，我们分析 μ_3 的变化如何影响性能，其中 μ_3^{**}表示公式（8）左右两侧相等时的取值。结果证实，理论预测的 μ_3^{**}与应用 LocalScore 对性能不起作用时的现实世界取值高度一致。当 $\mu_3 < \mu_3^{**}$时，LocalScore 会降低验证性能；而当 $\mu_3 > \mu_3^{**}$时，则会提升性能。

图 9b 展示了公式（8）右侧与左侧之差如何影响 TAR@0.1%FAR，差异越大，通常带来的提升越显著。



(a) (b)

图 9. (a) μ_3 对 TAR@0.1%FAR 的影响。垂直虚线表示理论预测的 μ_3^{**}，水平线表示不使用 LocalScore 时的性能。(b) 使用 LocalScore 后，式（8）右侧与左侧之差对 TAR@0.1% FAR 提升幅度的影响。阴影区域表示 99% 置信区间。



(a) k的影响。 (b) σ的影响。

图 10. k 和 σ 对 TAR@0.1%FAR 的影响。(a) 中，水平线表示不使用 LocalScore 时的性能。阴影区域为 99%置信区间。

不过，使用 KNN 时，验证性能略有下降，这可能是因为图 μ_3 中的实际值与理论值偏差较大；图 9a预测此处不会有任何提升。这可能是因为，与开放集生物特征识别相比，公式（8）中的正态性假设对验证的影响更大；图 10也观察到了这一现象，其中 KNN 往往会降低性能。尽管如此，LocalScore 仍能持续提升真实世界生物特征数据集上的验证性能，如图 7所示；除 BRIAR 上的步态识别外，均取得了显著提升。这表明，尽管公式（7）为真实世界场景提供了实用指导，但公式（8）在预测 LocalScore 带来的性能提升方面效果较差。

D.2. 额外消融研究

图10展示了在合成数据集上，LocalScore 对验证性能的影响。图10a中，随着 k 增大，TAR@0.1%FAR 逐渐下降，但与图4(a)中 FNIR@1%FPIR 大幅提升（达到约 15%）相比，下降幅度极小，最差约为 2%。这表明 LocalScore 在随机生成的特征上显著提升了开放集性能，而验证性能仅略有降低。

图 10b 考察了 σ（其决定类内变化）对验证性能的影响。随着 σ 增大，类内变化随之增大，而使用 LocalScore 与不使用 LocalScore 之间的性能差距则缩小。值得注意的是，即使在较高的 σ下，我们的方法仍能达到与基线相当的验证性能。这表明，LocalScore 在提升开放集性能的同时不会牺牲验证性能，从而为生物识别系统提供均衡而全面的评估。

E. 局限性

LocalScore 假设图库媒体分布良好，能够合理近似实际的图库特征分布。尽管这一假设在真实场景中未必总是成立，我们提出了一个理论框架（定理1），用于预测 LocalScore 是否会提升验证性能和开放集生物识别性能。我们在真实世界数据集以及随机生成的特征和分数上验证了这一框架。在实际应用中，部署前可以使用一个小型测试集评估 LocalScore 潜在的性能增益。

我们的评分算法需要为每个媒体特征提供额外存储，但由于若干因素，这一需求是可控的。生物特征通常较为紧凑，而大容量存储也日益经济实惠。此外，对图库样本进行聚类可以在不显著牺牲性能的情况下降低存储负载，并且云存储在处理数据时具有可扩展性和成本效益，不受本地限制。

F. 潜在社会影响

尽管我们的工作推动了生物识别技术的发展，并可能在安全和法证方面带来益处，但它也引发了对隐私和个人自由的担忧。敏感的生物识别数据（如面部特征）可能被滥用于未经授权的监控或用户画像，从而严重侵蚀公众信任并侵犯权利。我们优先使用经知情同意收集的数据，并避免直接收集任何生物识别数据。我们敦促社区采取透明且负责任的做法，以确保生物特征识别以积极方式服务社会，同时尽量减少潜在危害。