关于文本数量对作者检索的影响

Marco Peer[1,2], Robert Sablatnig[2], and Florian Kleber[2]

[1] iCoSys Institute, University of Applied Sciences and Arts Western Switzerland
[2] Computer Vision Lab, TU Wien, Austria

Abstract. 本文研究了作家检索任务，该任务根据手写相似性识别数据集中由同一作者撰写的文档。虽然现有的数据集和方法主要关注于页面级别的检索，我们通过评估行和词级别的检索来探索文本数量对作家检索性能的影响。我们考察了三种最先进的作家检索系统，包括手工制作的方法和基于深度学习的方法，并分析了它们在不同文本量下的性能。我们在 CVL 和 IAM 数据集上的实验表明，当仅用一行文本作为查询和库时，性能下降 20-30 %，但当至少包含四行时，检索准确率仍保持在全页性能的 90 % 以上。我们进一步表明，在低文本场景中，文本依赖的检索可以保持良好的性能。我们的研究结果还突出了手工制作特征在低文本场景中的局限性，像 NetVLAD 这样的深度学习方法优于传统的 VLAD 编码。

1 介绍

作者检索 (信息检索) 涉及通过分析笔迹相似性来识别数据集中由同一作者撰写的文档 [14] 。这一技术在数字人文和法医学中尤为重要，有助于识别未知文档的作者或追踪历史时期中的个人和社会群体 [9] 。信息检索 的工作原理是利用查询文档并根据其笔迹相似性对数据库中的其他文档进行排序，如图 1 所示。

尽管存在几个同时适用于现代和历史文档的数据集，如 CVL [15] 、IAM [19] 、ICDAR2013 [17] 、Historical-WI [10] 和 HisIR19 [9] ，这些数据集及其评估基准主要关注页面级别的检索，即对整个页面的特征进行编码和聚合。对于现代数据集，文献中提出的方法 [23,26] 报道了对整页文档超过 95 % 平均平均精度 (平均平均精度) 的表现。相比之下，本文通过以下问题: 当前 信息检索 方法需要多少最少文本量才能实现与整页检索相当的性能? 据我们所知，之前的研究尚未探讨 1) 文本数量对 信息检索 的影响以及 2) 在行或字级别对 信息检索 的评估。虽然历史数据集可能缺乏某些作者的足够手写样本（尤其是在分开的部分中，如 HisFrag20 [28] ，且未提供任何版面信息），现代数据集通常每个样本包括多行。因此，我们评估了三种最先进的 信息检索 系统（包括手工制作的和基于深度学习的方法），其中由一个特征提取器和一个基于码本的编码阶段组成，以评估行和字级别的检索性能的影响。我们的实验还考察了每个文档中行数的变化如何影响性能，揭示了文本数量的角色。此外，我们探讨当前方法是否能够在字级别捕捉到特定作者的特征，仅聚合来自单个字的特征。

我们的结果表明，在短查询和图库设置中，两个数据集的 平均平均精度 性表现下降约 20-30 %，其中每个样本仅使用一行文本。然而，当包括至少四行时，页面级别的准确性仍然保持在 90 % 以上。然而，我们还证明，当查询样本包含有限的文本而图库由完整页面组成时，检索仍然是有效的。在 Norhandv2 [29] 上进行评估时，这些结果也成立，这是一个由挪威手写文本组成的历史数据集。在词级别 信息检索 上，性能有所下降（最佳方法实现了约 13 % 的 平均平均精度

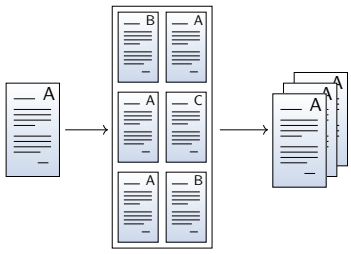


Fig. 1: 信息检索 的定义。文档（画廊）根据与查询的相似性进行排名，相关文档出现在排名列表的顶部。虽然 信息检索 的方法通常在页面级别进行评估，但本文侧重于行级和词级检索，以研究文本数量的影响。

），但我们展示了在这些低文本场景中，文本相关的检索仍能识别相似性。从方法上看，我们观察到手工特征在文本稀缺情况下表现困难，而 NetVLAD 在行级和词级实验中均优于传统的 局部聚合描述符向量 (向量量化) 编码。这些发现对于开发和基准测试未来的 信息检索 方法具有重要价值，特别是对于历史文献研究，其中来自特定作者的手写样本通常稀缺。

总之，我们的贡献是：

- 特征采样对检索性能影响的研究
- 详细基于行的评估三种最新的方法以评估文本数量对检索性能的影响，以及
- 对这些方法进行基于词的评价，以分析提取特征中的写作风格和词汇特定特征。

本文的其余部分组织如下：第 2 节讨论了 信息检索 方面的相关工作。第 3 节概述了本研究中使用的前沿 信息检索 方法以及我们的实验数据集。结果在第 4 节中展示，然后是第 **??** 节中的结论。

2 相关工作

下面我们概述了关于 信息检索 的相关工作，接着简要总结了对历史文献的类似研究。Fiel 和 Sablatnig [11] 提出了基于深度学习的 信息检索 的第一种方法，他们提出了一种 CaffeNet，一种 卷积神经网络 (卷积神经网络) ，利用分割的文本行进行训练，以通过倒数第二层激活的均值池化来构建页面描述符。在 [8] 中，Christlein 等人训练了一个 卷积神经网络 ，但使用 VLAD 和三角嵌入聚合特征。Christlein 等人。[6] 从图像块中提取局部特征，使用一种在以手写轮廓为中心的块上训练的 卷积神经网络 结构。从倒数第二层的激活函数导出的嵌入与 GMM 超级向量编码，随后进行 ZCA-白化规范化处理。与 SIFT 或神经网络不同，Christlein 等人。[5] 使用 向量量化 编码的轮廓 Zernike 矩来进行检索，这是一种手工设计的局部形状描述子。Keglevic 等人。[14] 提出了一种三重卷积神经网络 方法，其中以 尺度不变特征变换 (尺度不变特征变换) [18] 关键

点为中心的图像块用于训练一个基于相似性度量的三分支网络。网络的最终线性层用作嵌入，使用不同词汇表大小的多个 向量量化 [2] 编码进行编码。与大多数基于 深度学习 的方法从头开始训练 卷积神经网络s 不同，Liang 等人。[16] 利用迁移学习，通过结合 向量量化 [2] 使用预训练的 ResNet-50 [12] 进行编码。在 [26] 中，Rasoulzadeh 和 Babaali 采用 NetVLAD [1] ，这是一种 向量量化 编码的学习变体。他们的方法目前在如 CVL [15] 或 ICDAR2013 [17] 等当代数据集上处于最先进水平。此外，Peer 等人。 [23] 研究了 NetMVLAD，这是一种在 NetVLAD 层中具有多个词汇表的扩展版本，性能略优于 NetVLAD。

对于历史文献，Christlein 等人。[7] 提出了一个无监督的方法，该方法利用基于 ResNet20 的 VLAD 编码，应用于在 SIFT 关键点位置提取的图块。其目标标签是通过对相应的 SIFT 描述符进行聚类生成的。这种训练策略也被名为 HisFrag20 [28] 的相应数据集比赛的获胜者采用，该数据集包含人造的中世纪手稿片段。这些样本包含较少的文字，通常只有几个单词。Peer 等人。 [24] 表明，通过自监督的预训练，当样本文字较少时，信息检索 的性能有所提高，但比起完整页面的数据集，性能得分较低（未经预训练情况下为 46.6 % 平均平均精度 [21,24] ）。Raven 等 [27] 通过对基于 Transformer 的架构结合 向量量化 进行自监督训练，在两个包含完整页面的历史数据集上实现了最新的性能。Mattick 等 [20] 研究了在 Historical-WI [10] 上使用学习得来的池化来执行 信息检索 任务，显示传统的求和池化更为优越。

虽然本节中描述的方法都是文本无关的，但第一次尝试通过应用字符和 n -gram 方案实现文本相关的 信息检索 的方法是由 Peer 等人。[25] 在希腊纸草文档领域引入的。他们研究了仅聚合某些特定字符（如三元组 kai，即希腊文中的“和”）时的性能表现。然而，这一发现仅限于希腊文本，尚未进行关于文本数量对历史或现代数据集影响的一般性实验。

3 方法论

我们对 信息检索 使用三种最新的方法，即

- SIFT + 向量量化 ，
- ResNet20 + 向量量化
- 和 ResNet20 + NetVLAD [1] 。

这些方法的选择旨在涵盖各自 信息检索 管道的不同方面：特征提取阶段要么基于手工特征（如 SIFT），要么基于深度学习（如 ResNet20）。关于编码阶段，我们依赖于外部训练的码本（向量量化 ）或与特征提取器端到端联合训练的编码（NetVLAD）。虽然所有方法都基于 向量量化 ，如 Section 2 中所述，相关工作中提出的方法——特别是那些在 信息检索 [5,26] 上表现领先的——都是基于这种编码。

不同于依赖关键点检测算法，通常是 尺度不变特征变换 [7,23,26] ，我们直接在二值化图像的轮廓处（黑色像素代表墨水）采样关键点，因为这可以提高性能，特别是在手写稀少时 [4] 。利用这些关键点，我们计算 SIFT 描述符并提取 32×32 图块，然后将其传递给神经网络。图 2 显示了图块采样的示例。空白图块 - 不包含手写内容 - 将被丢弃。

在下面，我们将简要解释在我们的工作中使用的两个特征描述符。

Fig. 2: 32×32 补丁采样。

对于 SIFT 描述符，我们使用 RootSIFT 描述符，这是一种标准 SIFT 描述符的修改版本，旨在改善 Arandjelovic 等人提出的特征匹配。它由 l_1 归一化、平方根变换（L1 归一化向量的每个元素通过取平方根进行变换），然后进行 l_2 归一化组成。

作为一种基于深度学习的特征提取器，ResNet20 [12] 在相关工作中被应用 [7,26] 。该网络在手写轮廓处采样的 32×32 小块上训练。用于分类的全连接层被截断，且通过半困难三元组的三元组损失函数 [3] 直接训练嵌入。以作者标签为目标。

3.1 编码

信息检索 主要由基于 向量量化 的编码主导,无论是原始版本 [2] 还是可学习版本 NetVLAD [1] 。向量量化 通过将词汇聚类来获得 N_c 个聚类 $\{\mathbf{c}_0, \mathbf{c}_1, \ldots, \mathbf{c}_{N_c-1}\}$ ，并通过编码一组局部描述符 $\mathbf{x}_i$, $i \in \{0, \ldots, N-1\}$ 。

$$\mathbf{v}_k = \sum_{i=0}^{N-1} \mathbf{v}_{k,i} = \sum_{i=0}^{N-1} \alpha_k(\mathbf{x}_i)(\mathbf{x}_i - \mathbf{c}_k), \quad k \in \{0, \ldots, N_c - 1\}, \tag{1}$$

如果 $\mathbf{c}_k$ 是距离 $\mathbf{x}_i$ 最近的聚类中心，则为 $\alpha_k = 1$ ，否则为 0，因此使得原始的 VLAD 编码不可微分。然后通过连接向量 $\mathbf{v}_k$ 来获得最终的全局描述符。

对于 NetVLAD，Arandjelovi等人 [1] 建议使用一个可学习的层来解决在（1）中 α_k 的不可微分性，它通过为每个聚类中心 $\mathbf{c}_k$ 引入一个带有参数 $\{\mathbf{w}_k, b_k\}$ 的卷积层来学习软分配

$$\overline{\alpha}_k(\mathbf{x}_i) = \frac{e^{\mathbf{w}_k^{\mathrm{T}}\mathbf{x}_i + b_k}}{\sum_{k'} e^{\mathbf{w}_{k'}^{\mathrm{T}}\mathbf{x}_i + b_{k'}}}. \tag{2}$$

NetVLAD 的聚类中心 $\mathbf{c}_k$ 也在训练期间学习。

我们使用 l_2 归一化并接着进行求和池化，将一个实体（页面、行或级别）的所有嵌入 $\{\mathbf{V}_0, \mathbf{V}_1, \ldots, \mathbf{V}_{N_f-1}\}$ 进行聚合

3.2 全局描述符和检索

。

$$\mathbf{V} = \sum_{i=0}^{N_f-1} \mathbf{V}_i \tag{3}$$

获取全局描述符 $\mathbf{V}$ 。此外，为了减少视觉突发性 [13] ，我们应用功率归一化 $f(x) = \text{sign}(x)\sqrt{|x|}$ ，然后是 l_2 归一化。最后，通过 PCA 进行白化的降维。对于排序列表，通过计算全局描述符的余弦相似性来对实体进行排序。

3.3 数据集

我们的实验是在两个数据集上进行的，即 CVL 和 IAM 数据库，以下是简要描述，统计结果如表 1 所示。

Table 1: 所用评估数据集的统计信息。

	CVL	IAM
# of writers	283	483
# of documents	1409	1159
# of lines	11842	8979
# of words	87752	72939

CVL CVL [15] 数据集的训练集包含 27 位作者，他们贡献了七篇文本（其中一篇是德语，六篇是英语），共计 189 页，而测试集由 283 位作者组成，每人撰写五页（其中一页是德语，四页是英语）。

IAM IAM [19] 数据库由 657 位作者写的 1539 页组成。356 位作者只贡献了一页，其余 301 位作者写了两到 60 页。由于没有官方的测试划分，我们将所有贡献了两到三页的作者作为训练集（380 页），从而得到来自 483 位作者的 1159 页作为我们的测试集。

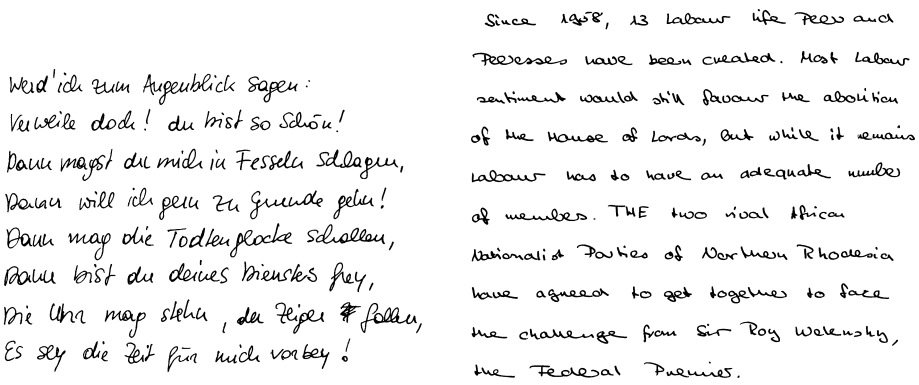
我们使用大津法对两个数据集进行二值化 [22] ，并使用提供的行和单词注释来进行我们的实验。对于 IAM 数据集，我们移除所有仅包含特殊字符的单词。两个数据集的示例显示在图 3 中。

4 结果

接下来，我们介绍我们的评估协议，并提供 信息检索 在页面级、行级和词级这三个粒度上的结果。

4.1 评估协议

所有方法都使用相同的协议进行评估。ResNets 在训练时三元组损失的边距为 $m = 0.1$ ，（Net-）向量量化 词汇的大小固定为 100，全球描述符进行白化并降至 256 维，以确保公平比较。训练在相应的训练集上进行，如小节 3.3 中所定义。我们评估使用的主要指标是 平均平均精度 ：对于长度为 Q 的查询列表中的每个 q ，平均准确率由以下方式给出：



(a) (b)

Fig. 3: （a ）CVL（ID: 0074-6）和（b ）IAM（ID: a01-007u）数据集的手写样本。

$$\mathrm{AP_q} = \frac{1}{R}\sum_{k=1}^{Q} \mathrm{Pr_q}(k) \cdot \mathrm{rel_q}(k), \tag{4}$$

其中 $\mathrm{Pr_q}(k)$ 是查询 q 的前 k 个元素中相关文档的比例，R 是图库中相关文档的总数，$\mathrm{rel_q}(k)$ 定义为

$$\mathrm{rel_q}(k) = \begin{cases} 1 & \text{if Label(q)} = \text{Label}(k) \\ 0 & \text{else} \end{cases}. \tag{5}$$

除非另有说明，用于评估的标签是作者标签。其次，我们使用（严格）Top-x 指标进行评估，该指标表明前 x 篇文档是否由同一作者撰写。

Table 2: 信息检索 在页面级别上的表现，每行使用 5k 特征。

	CVL		IAM	
	平均平均精度	Top-1	平均平均精度	Top-1
SIFT + 向量量化	96.1	98.3	92.5	99.4
ResNet20 + 向量量化	97.4	99.2	93.8	99.5
ResNet20 + NetVLAD	97.0	98.9	92.2	99.4

4.2 页面级检索

为了证明所选方法的有效性，我们评估了页面检索，其结果如表 2 所示。表现最佳的方法是 ResNet20，在两个数据集上达到 向量量化 ，尽管所有三种方法

的性能相似，差距在 1.3 个百分点内。通过对每行采样 5000 个特征，我们大约获得每个文档 40000 个特征。为了公平比较，我们在图 4 中评估了在两个数据集中每行使用的特征数量。我们评估了从 10 到 5000 个特征（包括 SIFT 描述符或补丁嵌入）的范围，并观察到以 ResNet 为基础的特征提取在每行约 1000 个特征时性能趋于饱和，而 SIFT 则在 5000 个特征时继续提高，这使得基于深度学习的方法在计算上更为高效。

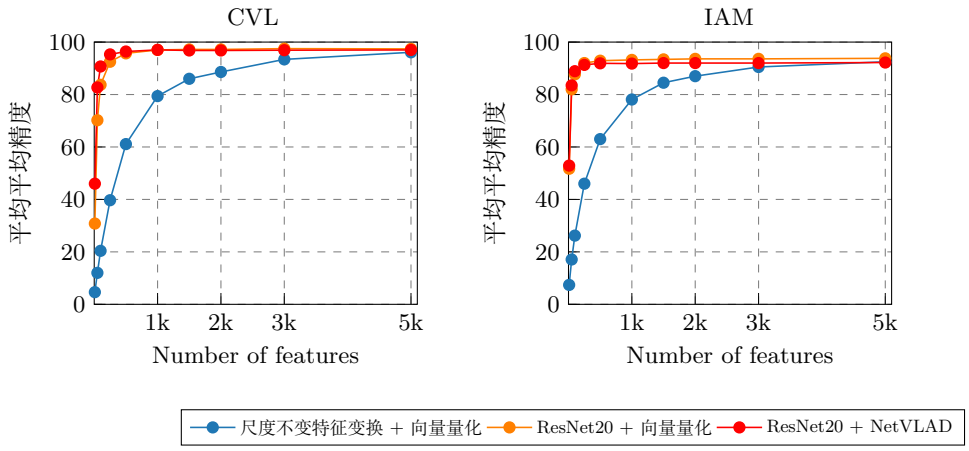


Fig. 4: 页面级检索：每行使用的特征数量对 CVL 和 IAM 数据集的 信息检索 性能的影响。

4.3 行级检索

本节中的实验是在行级别进行的，其中每一行用作查询，其余行则被排序。与页面级别 信息检索 相比，NetVLAD 表现优于 向量量化 ，如表 3 所示（对于 CVL 和 IAM 分别提高了 7.7 和 0.3 个百分点 平均平均精度 ）。然而，与子节 4.2 中的页面级检索结果相比，性能显著下降，下降了 30/20 个百分点 平均平均精度 。这表明这三种方法都受到编码文本减少的影响，IAM 行的表现略优于 CVL，这可能是因为 IAM 行平均包含更多的文本（每行 ≈ 7.4 与 ≈ 8.1 个单词）。

Table 3: 每行使用 5000 个特征的行级别 信息检索 。

	CVL		IAM	
	平均平均精度	Top-1	平均平均精度	Top-1
SIFT + 向量量化	47.3	82.4	55.8	85.6
ResNet20 + 向量量化	60.9	90.2	74.9	94.8
ResNet20 + NetVLAD	68.6	94.2	75.2	95.6

我们还研究了每行特征样本数量的影响，这对于行特别重要，因为文本的数量有限。结果如图 5 所示。传统的使用 SIFT 和 向量量化 的方法性能下降最为显著，即使每行有 5000 个特征，性能也未完全饱和。相比之下，对于 ResNet20（向量量化 和 NetVLAD），在每行仅使用 1000 个特征时，平均平均精度 分数就已经饱和。

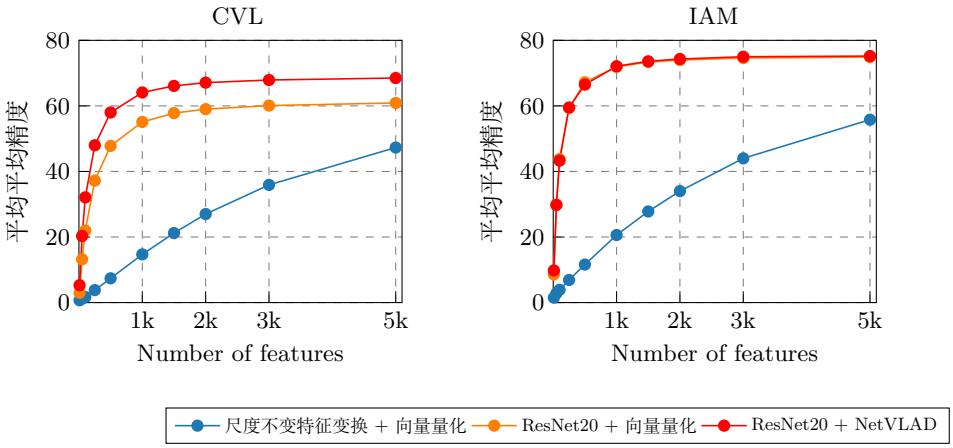


Fig. 5: 行级检索：特征数对 CVL 和 IAM 数据集每行的影响。

接下来，我们通过将查询文档的文本减少为一行或半页来关注检索任务，并将其与页面级表现进行比较。这种设置反映了领域专家的任务，例如那些在法医学领域的专家，他们可能拥有参考文档的数据库，但只有一个来自未知作者的小的手写样本。结果在表 4 中展示。对于单行和半页的实验，我们报告平均平均精度 的平均值，因为检索是多次进行的（每行一次并且每页的每部分两次）。有趣的是，我们观察到性能没有显著下降，这表明仅用一行文字作为查询来检索相关文档并不会对检索过程产生负面影响。

Table 4: 当使用一行（1L）、半页（HP）和整页（FP）作为查询时的 mAP。

	CVL			IAM		
	1L	HP	FP	1L	HP	FP
SIFT + 向量量化	95.9	96.0	96.1	92.5	92.5	92.5
ResNet20 + 向量量化	97.4	97.4	97.4	93.7	93.7	93.8
ResNet20 + NetVLAD	96.9	96.9	97.0	92.2	92.2	92.2

限制数据库中的文本数量（短查询 - 短库） 由于我们观察到行级别的 信息检索得分较低（表 3），我们现在旨在研究需要多少文本才能达到与页面级别检索

性能相当的水平。为了解决这个问题，我们通过逐渐合并文档的行来评估方法，从短查询和图库中的短样本开始。我们从单行开始，将数据库有效地拆分为各个单行，然后根据提供的行顺序逐步叠加每个文档的行。例如，当叠加两行时，将合并第 1 行和第 2 行，然后是第 3 行和第 4 行，依此类推。为了确保每个样本包含恰好 n 行，任何剩余行或不构成完整组的行都将被排除在检索过程之外。

该实验的结果展示在图 6 中。在此图中，平均平均精度 分数相对于表 2 中报告的页面级性能进行归一化。此归一化允许直接比较不同合并配置与基线页面级检索之间的差异。通过逐步增加用于检索的文本量，我们可以确定性能开始接近页面级检索的阈值，从而提供关于实现类似检索准确性所需文本上下文量的见解。

我们观察到行数对性能有很强的影响，尤其是对于 SIFT，它只有在 CVL 数据集使用超过七行和 IAM 数据集使用八行以上时才能达到页面级别 平均平均精度 的 80 % 以上。相比之下，基于深度学习的方法在较少文本的情况下表现显著更好。然而，当在这两个数据集上使用三行或更少的行时，它们的性能仍然低于页面级别 平均平均精度 的 90 %。

基于这些观察，我们得出结论，要达到与整页检索相当的性能，所需的最少文本量大约是四行。此门槛专门适用于深度学习特征提取器，这些提取器在四行文本的情况下可达到页面级性能的 90 % 以上。相比之下，SIFT + 向量量化 方法的表现不佳，仅能在每个样本使用四行时达到页面级 平均平均精度 的约 65 %。随着行数的增加，所有方法的性能都会提升，最终达到页面级 平均平均精度 。

图 6 中的直方图展示了两个数据集中少于 n 行的文件所占的百分比。我们认为，在 平均平均精度 中观察到的改进是由于这些方法能够捕捉书写风格，而不是因为数据集被“恢复”。这表明，即使是有限的文本，深度学习方法在提取区分作者的特征方面是可取的，而传统方法如 SIFT + 向量量化 需要更多的文本才能达到可比较的性能。

基于历史数据的行级检索 为了展示这些观察结果能够推广到更复杂的手写体，我们在历史数据集 Norhandv2 [29] 上评估了三种 WR 方法，该数据集包含挪威手写体，书写者按 40/60 进行划分，因为它也提供了行分割。首先，如表 5 所示，各种方法的性能显著降低。然而，文本数量的影响依然存在，只是由于页面级别得分较差，低文本情况下性能的相对下降较小。

最后，我们的论文关注的是最细粒度，即单词。我们还提供了硬性 Top- x 标准，因为这些方法在 CVL 和 IAM 数据集上仅能达到 平均平均精度 高达 9.3/13.6 % 的值。

表 6 展示了采用硬性标准和 mAP 评估的方法在词级检索性能上的表现，每种方法每个词使用 500 个特征。在 CVL 数据集中，ResNet20 + NetVLAD 在所有度量指标上都获得了最高分，优于 SIFT + VLAD 和 ResNet20 + VLAD 配置。在 IAM 数据集中，ResNet20 + NetVLAD 在 Top- x 表现优异，而 ResNet20 + VLAD 在 mAP 上稍微超过它，得分为 13.6 % 对比 13.3 %。总的来说，这些结果表明词级检索比在更大实体上的检索更为复杂，正如较低的性能指标所示。然而，以深度学习为基础的方法（尤其是 ResNet20 + NetVLAD）比传统特征提取方法表现更优。

我们在图 8 中展示了使用 ResNet20 和 NetVLAD 在单词级别上对 信息检索 的一些定性结果。首先，查询单词经常从同一行或相邻行中检索单词（如查询 `0052-1-0-0` 或 `0095-1-3-2` 所示），这表明模型捕捉到了书写风格的上下文或时

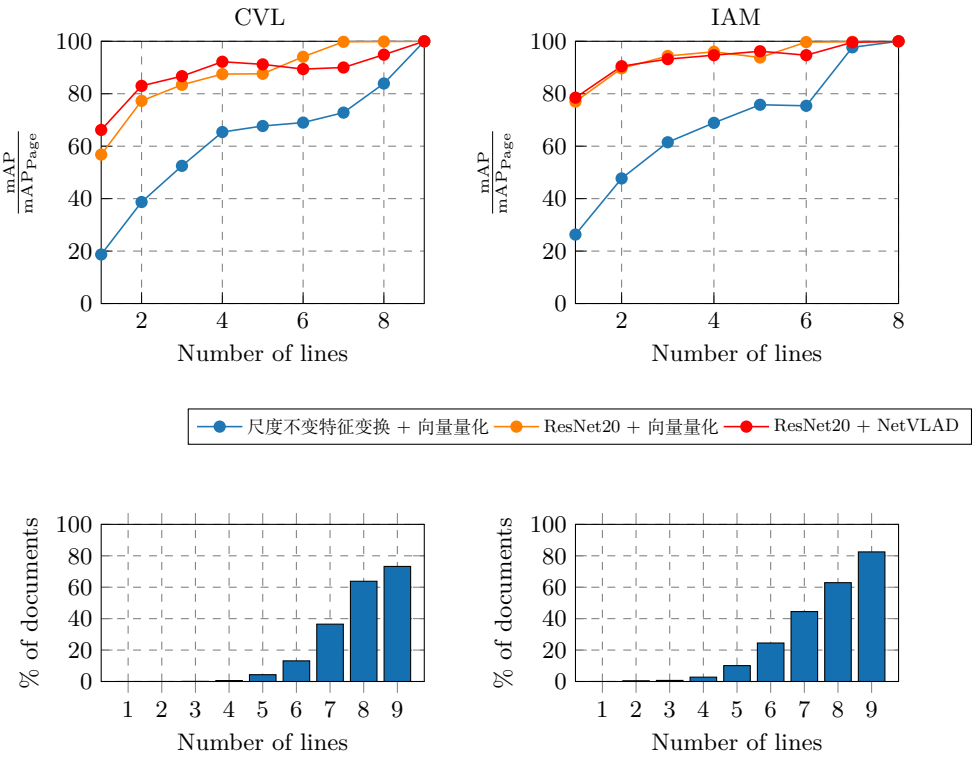


Fig. 6: 行级检索：对页面级检索 mAP_{Page} 的正常化 信息检索 表现。对于 CVL 和 IAM 数据集，我们逐渐合并文档的行数。

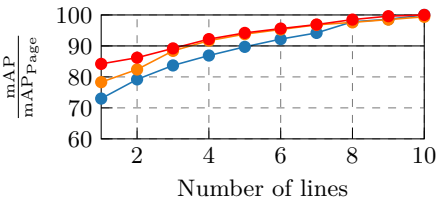


Fig. 7: Norhandv2: Line level retrieval

Table 5: Norhandv2: Full page retrieval performance

	mAP	Top-1
SIFT + 向量量化	21.1	78.6
ResNet20 + 向量量化	35.6	92.5
ResNet20 + NetVLAD	33.2	92.2

Table 6: 每个词使用 500 个特征在词汇层面评估的严格标准和 mAP。

(a) CVL

	Top-1	Top-3	Top-5	Top-10	mAP
SIFT + 向量量化	21.8	4.9	0.9	0.8	3.0
ResNet20 + 向量量化	35.0	12.7	7.1	3.0	7.5
ResNet20 + NetVLAD	49.3	22.8	13.9	6.2	9.3

(b) 身份认证模式

	Top-1	Top-3	Top-5	Top-10	mAP
SIFT + 向量量化	30.0	12.5	8.5	5.2	5.0
ResNet20 + 向量量化	47.2	27.4	20.5	13.5	13.6
ResNet20 + NetVLAD	54.7	32.7	24.2	15.2	13.3

间上的一致性。其次，我们看到模型的结果经常依赖于相似的单词，即使它们不是由同一个作者写的，例如对于 `0095-1-3-3` 的 “on”。对于其他单词，例如作者 0201 的 “fallen” 或 “Die”，我们检索到了具有相似异体字的单词（“D”、“e” 或 “l”、“f” 的圈）。

最后，我们引入了单词特定检索任务，其中数据库仅包含含有特定单词的样本，目标是检索由同一作者书写的样本。我们专注于评估与 NetVLAD 结合的 ResNet20，因为它提供了最佳性能。在图 9 中，我们展示了两个数据集中最常见的 20 个单词的 平均平均精度 表现。与完整的基于单词的 信息检索 相比，结果改善有两个主要原因：1）相关项目数量减少（将每个作者的样本限制为仅一个单词），以及 2）能够比较单词中特定字符的异形字体。例如，即使是单个字符 ‘a’，我们也实现了 29.3 % 和 36.2 % 的 平均平均精度 得分。在 CVL 数据集中，德语单词 ‘Dann’（然后）获得了最高的 平均平均精度 ，即 71.4 %，这可能是由于它在同一页上的重复出现。然而，我们没有观察到任何明确的模式表明更多字母的单词会带来更好的表现。我们得出结论，基于相同单词实例的检索对于当前 信息检索 方法更为有利——尤其是在可用的手写文本有限时——即便方法是在与文本无关的方式下训练的。

在这篇论文中，我们为当代手写检索领域中的三种最先进的方法（信息检索）引入了新的基准测试，特别是在使用 CVL 和 IAM 数据集的情况下。我们展示了在页面、行和单词层次的结果，分析了全局描述符的特征样本大小和文本数量的影响，特别关注于行检索。研究结果表明，尽管这些方法在页面层次上实现了超过 95 % 平均平均精度 的准确率，但在只有一行文本可用时，其性能下降了大约三分之一。此外，我们的实验表明，随着行数的增加，检索性能有所改善，四行文本的检索性能接近页面层次检索准确率的 90 %。然而，在单词层次检索方面，当前方法表现不佳。我们还证明了作家特定单词检索（数据库中仅包含特定单词）比包含多个不同单词的数据库获得更好的结果。

对于未来的工作，我们建议在低文本场景中评估 信息检索 方法，以更清楚地了解当前算法的局限性。我们还建议，当每个作者可用的手写文本有限时，如在历史文档的情况下，比较相同单词实例可以提高性能，因此依赖文本的算法

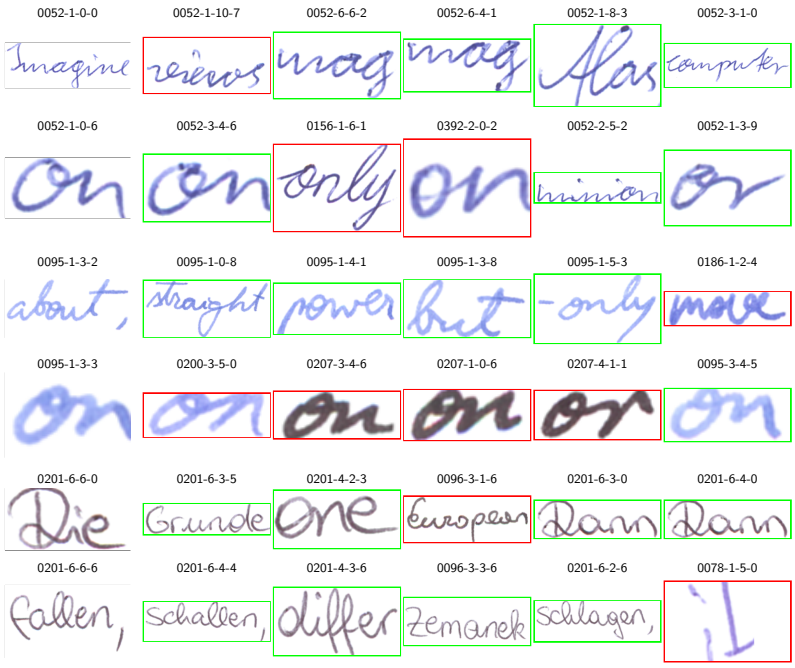
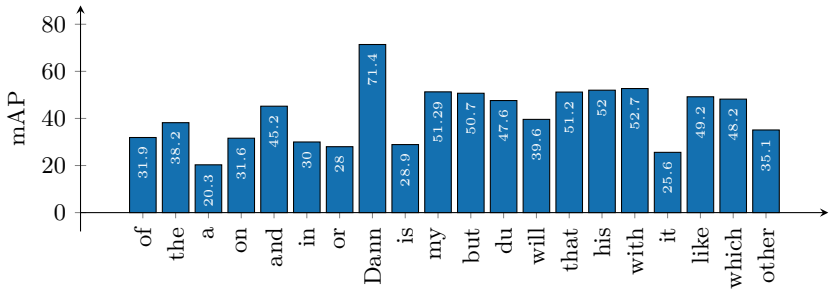
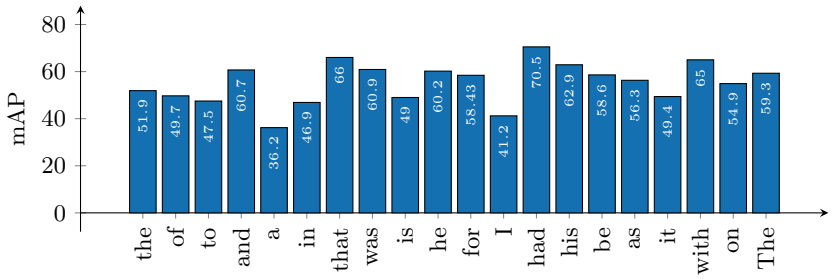


Fig. 8: 使用 ResNet20 + NetVLAD 对 CVL 数据集进行单词检索的定性结果。左侧图像是查询图像，后面是五个最相似的单词（ID：`Writer-Page-Line-Word`）。



(a) CVL



(b) IAM

Fig. 9: 在相应数据集中最常见的 20 个词的 信息检索 表现。

可能更有效。最后，我们推荐探索行或单词级别的检索，以促进开发更高级的特征聚合方法（例如，学习变体而不是简单的求和池化），因为当可用特征集较小时，特征的聚合可能更具影响力。

References

1. Arandjelovic, R., Gronát, P., Torii, A., Pajdla, T., Sivic, J.: Netvlad: CNN architecture for weakly supervised place recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 5297–5307 (2016)
2. Arandjelovic, R., Zisserman, A.: All About VLAD. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1578–1585. Portland, OR, USA (2013)
3. Balntas, V., Johns, E., Tang, L., Mikolajczyk, K.: PN-Net: Conjoined Triple Deep Network for Learning Local Image Descriptors. arXiv:1601.05030 [cs] (Jan 2016), arXiv: 1601.05030
4. Christlein, V.: Handwriting Analysis with Focus on Writer Identification and Writer Retrieval. Ph.D. thesis (2018)
5. Christlein, V., Bernecker, D., Angelopoulou, E.: Writer identification using VLAD encoded contour-zernike moments. In: 13th International Conference on Document Analysis and Recognition, ICDAR 2015, Nancy, France, August 23-26, 2015. pp. 906–910 (2015)

6. Christlein, V., Bernecker, D., Maier, A., Angelopoulou, E.: Offline Writer Identification Using Convolutional Neural Network Activation Features. In: Pattern Recognition. vol. 9358, pp. 540–552 (2015)
7. Christlein, V., Gropp, M., Fiel, S., Maier, A.K.: Unsupervised feature learning for writer identification and writer retrieval. In: 14th IAPR International Conference on Document Analysis and Recognition, ICDAR 2017, Kyoto, Japan, November 9-15, 2017. pp. 991–997 (2017)
8. Christlein, V., Maier, A.K.: Encoding CNN activations for writer recognition. In: 13th IAPR International Workshop on Document Analysis Systems, DAS 2018, Vienna, Austria, April 24-27, 2018. pp. 169–174 (2018)
9. Christlein, V., Nicolaou, A., Seuret, M., Stutzmann, D., Maier, A.: ICDAR 2019 Competition on Image Retrieval for Historical Handwritten Documents. In: 2019 International Conference on Document Analysis and Recognition (ICDAR). pp. 1505–1509. Sydney, Australia (2019)
10. Fiel, S., Kleber, F., Diem, M., Christlein, V., Louloudis, G., Nikos, S., Gatos, B.: ICDAR2017 Competition on Historical Document Writer Identification (Historical-WI). In: 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). pp. 1377–1382. Kyoto (2017)
11. Fiel, S., Sablatnig, R.: Writer Identification and Retrieval Using a Convolutional Neural Network. In: Computer Analysis of Images and Patterns. vol. 9257, pp. 26–37. Springer International Publishing (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778. IEEE, Las Vegas, NV, USA (2016)
13. Jégou, H., Douze, M., Schmid, C.: On the burstiness of visual elements. In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA. pp. 1169–1176 (2009)
14. Keglevic, M., Fiel, S., Sablatnig, R.: Learning Features for Writer Retrieval and Identification using Triplet CNNs. In: 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR). pp. 211–216. Niagara Falls, NY, USA (2018)
15. Kleber, F., Fiel, S., Diem, M., Sablatnig, R.: CVL-DataBase: An off-line database for writer retrieval, writer identification and word spotting. In: 12th International Conference on Document Analysis and Recognition, ICDAR 2013, Washington, DC, USA, August 25-28, 2013. pp. 560–564 (2013)
16. Liang, D., Wu, M., Hu, Y.: Offline Writer Identification Using Convolutional Neural Network and VLAD Descriptors. In: Artificial Intelligence and Security, vol. 12736, pp. 253–264 (2021)
17. Louloudis, G., Gatos, B., Stamatopoulos, N., Papandreou, A.: ICDAR 2013 Competition on Writer Identification. In: 2013 12th International Conference on Document Analysis and Recognition. pp. 1397–1401. Washington, DC, USA (2013)
18. Lowe, D.G.: Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
19. Marti, U.V., Bunke, H.: The IAM-database: an English sentence database for offline handwriting recognition. International Journal on Document Analysis and Recognition **5**(1), 39–46 (2002)
20. Mattick, A., Mayr, M., Seuret, M., Kordon, F., Wu, F., Christlein, V.: Evaluating learned feature aggregators for writer retrieval. International Journal on Document Analysis and Recognition (IJDAR) **27**(3), 265–274 (Jun 2024)

21. Ngo, T.T., Nguyen, H.T., Nakagawa, M.: A-VLAD: an end-to-end attention-based neural network for writer identification in historical documents. In: 16th International Conference on Document Analysis and Recognition, ICDAR 2021, Lausanne, Switzerland, September 5-10, 2021, Proceedings, Part II. vol. 12822, pp. 396–409 (2021)
22. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics **9**(1), 62–66 (1979)
23. Peer, M., Kleber, F., Sablatnig, R.: Writer retrieval using compact convolutional transformers and netmvlad. In: 26th International Conference on Pattern Recognition, ICPR 2022, Montreal, QC, Canada, August 21-25, 2022. pp. 1571–1578 (2022)
24. Peer, M., Kleber, F., Sablatnig, R.: SAGHOG: Self-Supervised Autoencoder for Generating HOG Features for Writer Retrieval. In: Document Analysis and Recognition - ICDAR 2024. pp. 121–138. Athens, Greece (2024)
25. Peer, M., Sablatnig, R., Serbaeva, O., Marthot-Santaniello, I.: Kairacters: Character-level-based writer retrieval for greek papyri. In: Pattern Recognition. pp. 73–88 (2025)
26. Rasoulzadeh, S., BabaAli, B.: Writer identification and writer retrieval based on netvlad with re-ranking. IET Biom. **11**(1), 10–22 (2022)
27. Raven, T., Matei, A., Fink, G.A.: Self-supervised Vision Transformers for Writer Retrieval, p. 380–396. Springer Nature Switzerland, Athens, Greece (2024)
28. Seuret, M., Nicolaou, A., Maier, A., Christlein, V., Stutzmann, D.: ICFHR 2020 competition on image retrieval for historical handwritten fragments. In: 17th International Conference on Frontiers in Handwriting Recognition, ICFHR 2020, Dortmund, Germany, September 8-10, 2020. pp. 216–221 (2020)
29. Tarride, S., Schneider, Y., Generali-Lince, M., Boillet, M., Abadie, B., Kermorvant, C.: Improving Automatic Text Recognition with Language Models in the PyLaia Open-Source Library. In: Document Analysis and Recognition - ICDAR 2024. p. 387–404 (2024)