组织病理图像中的精简深度卷积激活特征（R-DeCAF）可提升乳腺癌诊断的分类性能

Bahareh Morovati、Reza Lashgari、Mojtaba Hajihasani 和 Hasti Shabani *

伊朗德黑兰沙希德·贝赫什提大学医学科学与技术研究所* 通讯作者：ha_shabani@sbu.ac.ir

摘要：乳腺癌是全球女性中第二常见的癌症。病理学家诊断乳腺癌的过程既耗时，又具有主观性。计算机辅助诊断框架通过自动分类数据来减轻病理学家的工作负担，其中深度卷积神经网络（CNNs）是有效的解决方案。从预训练卷积神经网络的激活层提取的特征称为深度卷积激活特征（DeCAF）。本文分析发现，并非所有 DeCAF 特征都必然能提高分类任务的准确率，而降维起着重要作用。因此，我们采用不同的降维方法，通过提取 DeCAF 特征的本质，实现特征的有效组合。为此，我们提出了精简深度卷积激活特征（R-DeCAF）。在该框架中，预训练卷积神经网络（如 AlexNet、VGG-16 和 VGG-19）以迁移学习模式用作特征提取器。从上述卷积神经网络的第一个全连接层提取 DeCAF 特征，并使用支持向量机进行二分类。在非线性和线性降维算法中，主成分分析（PCA）等线性方法能够更好地组合深度特征；在考虑特征特定累积解释方差（CEV）的情况下，仅使用少量特征即可提升分类任务的准确率。我们使用实验 BreakHis 数据集验证了所提出的方法。全面的结果表明，分类准确率最高可提升 4.3%，同时计算时间更短。使用预训练 AlexNet 作为特征提取器、PCA 作为特征降维算法时，针对 400× data 获得的最佳准确率为 91.13%，特征向量大小（FVS）为 23，CEV 等于 0.15。

关键词： 乳腺癌；深度特征提取；特征降维；组织病理图像；预训练卷积神经网络

---

1. 引言

乳腺癌（BC）是全球女性死亡的主要原因之一，但男性也可能罹患乳腺癌。乳腺癌的诊断通常包括通过触诊进行初步检测，以及通过超声成像或乳腺X线摄影进行定期检查；而对可能存在的恶性组织生长的诊断则通过乳腺组织活检进行检测 [1]。根据世界卫生组织（WHO）的数据，乳腺癌正影响大量女性的健康 [2]。近期研究预测，到 2030 [3] 年，乳腺癌新增病例将达到约2,700万例。早期发现乳腺癌对于采取适当治疗和降低死亡率至关重要。然而，乳腺癌诊断可能不够准确，因为病理学家只能在显微镜下通过目视检查样本 [4, 5]。针对这些挑战，利用计算机辅助诊断和卷积神经网络（CNNs）进行图像分类的自动分类成为活跃的研究领域，旨在以更低的误诊概率实现精准诊断和快速检测。

当前关于乳腺癌检测的最先进研究证实，在分类任务中，卷积神经网络比传统手工特征更可靠、速度更快 [6]。然而，训练卷积神经网络可能需要更长时间，而且设计此类网络需要专业知识 [6-8]。文献中报告的一种可行解决方案称为深度卷积激活特征（DeCAF），也称为深度特征 [6,

8, 9]。这些方法重复利用预训练卷积神经网络来提取深度特征，并将其输入分类器以作出最终判定。

Filipczuk 等人研究了乳腺癌组织病理学数据集（737 张图像）中的手工特征 等人 [10] ，采用圆形霍夫变换将细胞核分割为圆形。他们使用 K近邻（KNN）作为分类器，取得了 98.51% 的最高准确率 [10]。然而，虚拟切片中的感兴趣区域并非自动选取，因此这一过程耗时。此外，该方法无法保证全局最优，而椭圆分割需要更精确的模型，计算开销也更大。在 Sharma 和 Mehra 的另一项研究中，他们从 BreakHis 数据集中提取了颜色、形状和纹理等手工特征，并将其输入支持向量机（SVM）和随机森林（RF）等传统分类器。他们报告称，具有 1000 棵决策树的 RF 在 40× 数据 [11]上可达到 90.33% 的准确率。此外，他们还比较了手工特征与深度特征。使用 VGG-16网络对深度特征进行分类时，40× 数据的准确率为 93.97%。他们指出，手工特征的性能并不理想，因为这需要深入了解癌细胞形态，而深度特征是更好的替代方案。Alhindi 等人针对 KIMIA Path960数据集，将局部二值模式（LBP）、方向梯度直方图（HOG）等手工特征与使用预训练 VGG-19 提取的深度特征进行了比较 [12]。采用 LBP 特征和 SVM分类器时，准确率最高，为 90.52%。值得一提的是，LBP 的特征向量大小为 1182，几乎是所提取深度特征特征向量大小的两倍。

为处理组织病理图像中的深度特征，Spanhole 等从预训练 AlexNet 的不同全连接层中提取 DeCAF，并结合逻辑回归分类器，利用 BreakHis数据集诊断乳腺癌 [6]。结果表明，迁移学习是一种可行的替代方案，在40× 数据上的准确率达到 84.6%。随后，Deniz 等开发了一种框架，以利用两个预训练卷积神经网络对 BreakHis数据集进行二分类。他们结合了 AlexNet 和 VGG-16 的 DeCAF 特征，随后使用 SVM分类器，达到了 84.87% 的准确率 [8]。在[13], Kumar 等提出了 VGG-16 的一种变体，其中移除了所有全连接层，并使用不同分类器对 CMT 和 BreakHis数据集进行了评估。对于来自 BreakHis数据集的200× 数据，报告的最佳准确率为 97.01%，其 FVS 为 1472。

为克服训练数据集不足的问题，研究人员将组织病理图像划分为互不重叠或随机的图像块，并将其作为预训练卷积神经网络的输入进行特征提取 [1, 6, 14]。然而，提取某些图像块可能会导致分类结果存在不确定性 [15]。为提高分类的准确率，一些方法着重于从头开始训练卷积神经网络，或对预训练卷积神经网络进行微调 [15-19]。其中一些方法取得了更高的性能，但需要耗费较长时间，并且必须精确调整超参数。在某些情况下，与迁移学习技术相比，训练模型或微调所有层可能无法取得更好的性能 [7, 16, 17]。此外，无论是面临缺乏训练数据集来训练深度模型的问题，还是仅添加少量训练数据以重新训练整个模型，迁移学习都能有效解决这一问题 [13, 14, 16]。

近年来，深度特征的降维或特征选择引起了研究人员的关注。Alinsaif 等应用无限潜在特征选择（I LFS）方法，从经过或未经过微调的预训练卷积神经网络（如 ResNet 和 DenseNet）中选择排名靠前的特征。据报道，在使用 SVM分类器对 BreakHis数据集进行二分类时，准确率为 97.96%，其中 FVS 为 1300 [16]。此外，Gupta 等提出使用极端梯度提升（XGBoost）来减少从 ResNet 中提取的特征数量，并采用信息论度量（ITS）选择最优层数。据报道，在40× 数据上，准确率为 97.07±1.18%，其中 FVS 为 500。尽管如此，随着特征数量减少，准确率有所下降 [20]。

本研究主要探讨降维对使用较少特征捕获信息丰富的特征所产生的影响。我们的分析表明，并非所有深度特征都必然能提高分类任务的准确率，而降维发挥着重要作用。我们提出了 R-DeCAF 特征，以较低的计算时间捕获数据的本质。为实现这一目标，将预训练卷积神经网络（AlexNet、VGG-16 和 VGG-19）的权重保持在冻结模式，并从第一个全连接层提取深度特征；该层的特征向量大小较高，即 FVS = 4096。为了减小提取的深度特征的大小，我们采用了不同的线性和

非线性降维方法，例如主成分分析、奇异值分解（SVD）、线性判别分析（LDA）、核主成分分析（kPCA）和 t分布随机邻域嵌入（t-SNE），以生成 R-DeCAF 特征。在 BreakHis数据集上，使用径向基函数核的支持向量机对 DeCAF 特征与 R-DeCAF 特征（通过线性方法降维）进行全面比较，结果表明，将特征数量控制在 120 个以下，不仅能提高分类准确率，还能缩短计算时间。

本文组织如下：第2节介绍组织病理学数据集；第3节阐述问题 formulation，并详细介绍所提出的用于提取 R-DeCAF 的模型。在第4节中，讨论实验案例研究以及所得结果的比较分析。最后，第5节给出结论和未来研究方向。

2. 组织病理学数据库

BreakHis 数据库 [21] 由一个实验室（病理解剖学与细胞病理学，巴拉那，巴西）开发，是本研究使用的可公开访问的乳腺癌组织病理学数据集。该选定数据集包含乳腺癌组织的显微组织病理图像，共有来自82名患者的7,909幅乳腺癌组织图像，这些图像以40×、 100×、 200× 和400× 放大倍数获取。该数据集包含2480个良性样本和5429个恶性样本，彩色图像尺寸为700×460。此外，良性和恶性肿瘤还被划分为若干亚组。该数据集中的样本通过开放式手术活检（SOB）方法采集，并采用苏木精-伊红方法染色。每个图像文件名中都包含有关图像的存储信息，例如活检操作方法、放大倍数、癌症类型及其亚型，以及患者标识信息。图1展示了该数据集中不同放大倍数下的良性和恶性肿瘤样本图像。

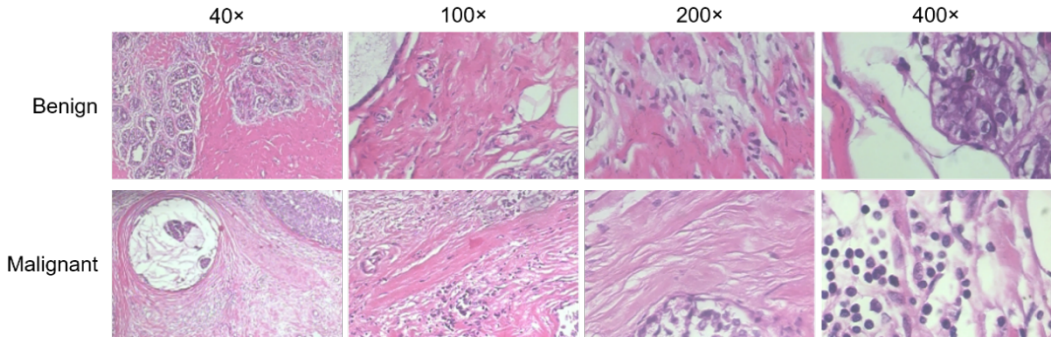


图 1. BreakHis 数据库中不同放大倍数下的样本图像。第一行来自良性肿瘤的同一张切片，第二行来自恶性肿瘤的同一张切片。

3. 方法

本研究从 BreakHis组织病理学数据集中进行深度特征提取，并表明从预训练卷积神经网络提取的所有特征并不能有效地对数据进行分类。因此，本研究的首要目标是降低所提取深度特征的特征向量大小，以保留信息丰富的特征并移除那些会造成误导或对分类不起重要作用的冗余特征。本研究考虑了多种用于从 BreakHis数据集中进行深度特征提取的当代预训练卷积神经网络模型。这些预训练卷积神经网络基于 ImageNet 数据集进行训练，该数据集包含超过 1400 万张自然图像，分为 1000 个类别。

迁移学习技术的优势在于，能够避免从头开始训练卷积神经网络所需的耗时过程 [6, 13, 16]。换言之，训练卷积神经网络和对其进行微调，分别需要海量数据和中等规模的数据。然而，迁移学习涉及不同的结构，且不需要海量数据。实际上，迁移学习是一种将一个领域中学习到的知识迁移到特定任务中的方法，旨在实现分类或特征提取。因此，目标是将从自然图像中学到的知识迁移到乳腺癌组织病理图像中，从而简化诊断

过程。由于预训练卷积神经网络是在包含大量类别和样本的大型数据集上训练的，因此能够提供规模庞大的网络，并通过有效组合特征来对数据进行分类 [6, 13]。

在迁移学习模式下，考虑冻结模式中的权重，从 AlexNet、VGG-16 和 VGG-19 网络的第一个全连接层中自动提取深度特征。因此，冻结基于 ImageNet 数据集训练的权重，可以使模型准备好使用所有已定义的预训练权重。之所以考虑 AlexNet、VGG-16 和 VGG-19 的第一个全连接层，即 FC6，是因为 FC6 层能够提供信息量更丰富的特征，从而实现更准确的分类 [6, 22]。

3.1. R-DeCAF 所提出的框架

我们通过应用适当的降维算法，将固定为 4096 的 DeCAF 特征的特征向量大小（FVS）降至低于 120，从而生成 R-DeCAF 特征。根据数据中的冗余或误导性特征 [23]，，并基于我们的分析，分类任务并不需要所有 DeCAF 特征。此外，高维特征向量会显著影响机器学习算法对数据进行拟合的性能，这通常被称为“维数灾难” [24]。因此，我们通过分析不同的降维技术，提出了 R-DeCAF 特征，以较低的计算时间捕捉数据的本质。所提出的框架的架构如图 2 所示。

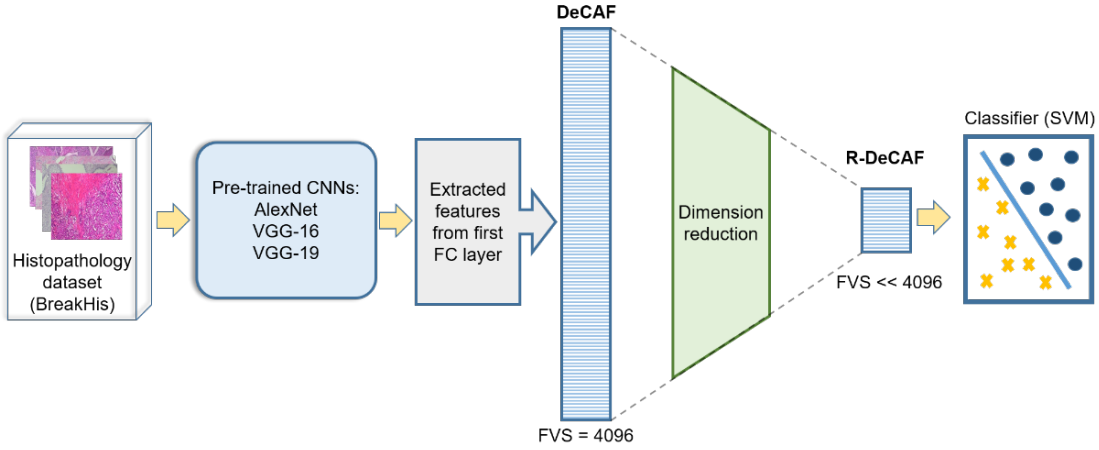


图 2. 所提出框架的示意图。

3.2. 预训练卷积神经网络

本文研究了 AlexNet、VGG-16 和 VGG-19 这三种预训练卷积神经网络，并将其作为所提出框架的基础，用于提取深度特征。AlexNet 被认为是 Krizhevsky 等人提出的首个深度卷积神经网络模型 [25]。该网络包含五个卷积层和三个全连接层，其中最后一层的神经元数量取决于数据的类别数。第一个和第二个全连接层均包含 4096 个神经元。Simonyan 等人在 2014 [26] 中提出了层数更多的 VGG-16 和 VGG-19 卷积神经网络。这两种卷积神经网络的所有层均使用尺寸为 3×3 的小型滤波器，以捕捉图像中的 精细细节并控制参数数量。VGG-19 有 19 个权重层，VGG-16 有 16 个权重层 [26]。需要说明的是，为了便于在本文中使用 PyTorch 库中的卷积神经网络模型，所有输入图像均调整为 224×224 。

3.3. 特征降维算法

本研究分析了针对 DeCAF 特征的不同降维方法，并将其分为两类：线性方法和非线性方法。前者包括 PCA、SVD 和 LDA，后者包括 kPCA 和 t-SNE [23, 27]。PCA 是一种线性无监督算法，可通过计算线性变换生成新特征。可以根据数据的协方差矩阵计算特征向量和特征值，

以确定数据的主成分（PC）。PCA 将数据中的最多信息保留在第一个 PC 中，并按递减顺序继续排列，因为主方向及其对应的 PC 被视为数据方差最大的方向 [23, 24]。SVD 是另一种适用于稀疏数据的线性降维方法。矩阵的 SVD 是将主矩阵分解为三个矩阵。该方法选取最大的奇异值，其特征值和特征向量与 PCA 方法一样按递减顺序排列。因此，输入矩阵将以低维形式重构 [23]。LDA 是另一种线性有监督降维方法，重点关注两个关键概念，即“类间散度”和“类内散度”。其主要目标是最大化“类间散度”或类别可分性。因此，LDA 可以选出能够以最佳方式区分数据类别的分量。需要指出的是，降维后的分量/特征数量可以等于或小于类别数量-1 [23, 28]。

在非线性降维方法中，核主成分分析是一种常见的无监督技术。当主成分分析方法效果不佳且数据结构呈非线性时，核主成分分析可能表现得更好。核主成分分析利用“核技巧”，在高维空间中降低原始数据的维度。在高维空间中，决策边界变为线性。在该方法中，根据按降序排列选出的降维特征向量，计算核矩阵的特征值和特征向量。将原始矩阵与特征向量相乘，以重建新的降维数据 [24, 27]。非线性无监督方法 t-SNE 是一种常见的数据探索与可视化技术。该方法将数据映射到低维空间，例如二维或三维空间。t-SNE 将数据点对之间的高维欧几里得距离 x_i, x_j 转换为条件概率 $p_{j|i}$，以表示数据点对之间的相似性；同时，它还将高维数据点 x_i, x_j 所对应的低维数据点 y_i, y_j 之间的相似性转换为条件概率。若数据点 y_i, y_j 能够在高维空间 [23, 29] 中反映数据点之间的相似性，则条件概率 $p_{j|i}$ 和 $q_{j|i}$ 将相等，该高维空间由 $q_{j|i}$定义。

3.4. 累积解释方差（*CEV*）与特征规模

为减少特征的维度，我们采用了特征值及其对应的累积解释方差（CEV）。实际上，确定主成分的最佳数量是一项具有挑战性且至关重要的工作，而 CEV 是解决这一难题的一种方法。CEV 是方差的累积，用于表示新特征（即主成分）的方差总和占累积方差的百分比，该百分比随主成分数量而变化 [23]。图 3 展示了从预训练 AlexNet、VGG-16 和 VGG-19 的第一个全连接层中提取的 DeCAF 特征的 CEV，这些特征对应于 BreakHis数据集的全部放大倍率数据。如图所示，当主成分数量超过约 2560 个时，CEV 的变化微乎其微。换言之，前 2560 个主成分涵盖了 100% 的累积方差，而前 512 个主成分涵盖了 67% 的数据方差。

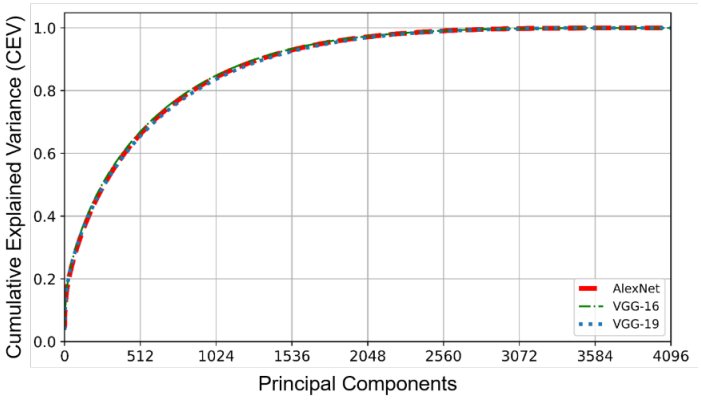


图 3. 使用预训练卷积神经网络对 BreakHis数据集的全部放大倍率数据提取的 DeCAF 特征的累积解释方差。

可以得出结论，按照经验法则，几乎一半经过变换的特征（即主成分）在分类中并不发挥重要作用。Zhong 等 [9] 采用了这一简单规则。主要原因是提取的深度特征之间存在高度相关性。为了进一步探究这一问题，我们考察了从 5% 到 100% 的完整 CEV 范围，其中 100% 表示使用了从 DeCAF 特征中获得的所有主成分，以评估分类准确率。此处的平均准确率取自三种预训练卷积神经网络，并基于将特征向量划分为训练集和测试集的 10 种不同方式计算。图 4 展示了 BreakHis数据集全部放大倍率数据（7909 张图像）的这些结果。

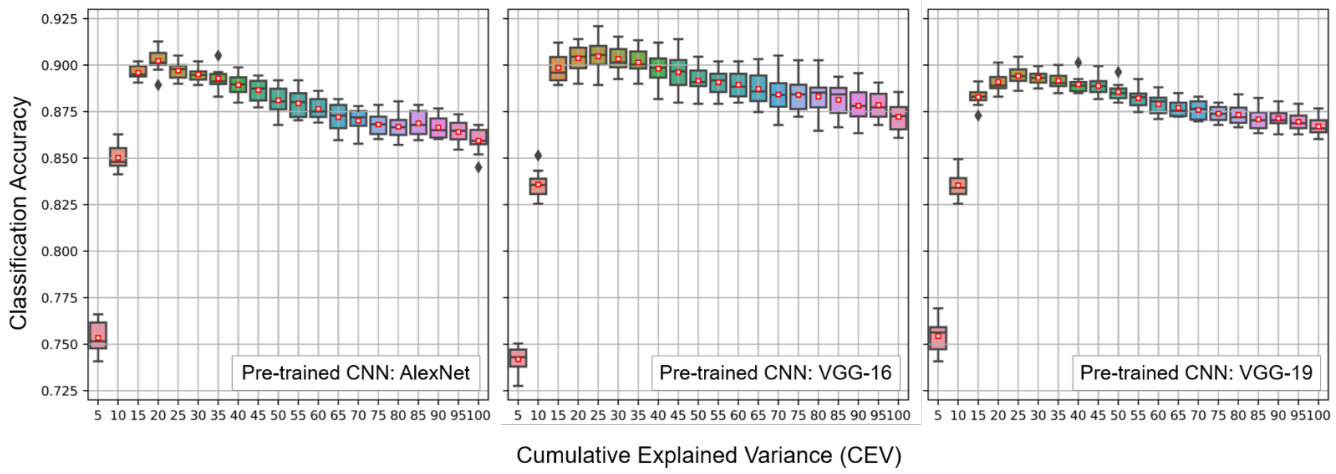


图 4. BreakHis数据集所有放大倍率数据的平均分类准确率与累积解释方差（CEV）的关系。

首先，这些结果证实，使用低于 100% 但高于 15% 的 CEV，不仅能保持相同的性能，还能提高深度特征的分类效果。其次，图 4 的结果进一步表明，与其使用大量特征，不如向分类器输入更有效、更合适的特征。因此，使用更少的特征也能获得更高的准确率，而特征的内容在分类任务中起着至关重要的作用。这些结果表明，考虑大量特征并不一定会带来更高的性能，而且从预训练 AlexNet、VGG-16 和 VGG-19 中提取的所有 DeCAF 特征在分类中并非都具有说服力和信息量。图 4 显示，仅保留 20% 至 25% 的 CEV，相比保留 50% 的 CEV 能带来更显著的提升。此外，对于预训练 AlexNet、VGG-16 和 VGG-19，特征向量大小也分别从 4096 大幅降至 63、103 和 93。关于这项研究的更多详情见表 1。

3.5. 分类器

此处选择支持向量机算法进行分类，因为它能够通过核技巧处理高维数据和非线性分类 [13]。该技术用于评估 DeCAF 特征和 R-DeCAF 特征在分类任务中的表现，以预测样本是良性还是恶性。采用径向基函数核的训练后支持向量机被视为常用的核函数；通过网格搜索比较不同核函数，并使用支持向量机参数 C=5。将定义的数据集划分为训练集（80%）和测试集（20%）。采用拆分方法，并通过对 10 次不同拆分的结果取平均值来报告结果。使用 PyTorch 和 Scikitlearn 库获得全面的结果，以验证所提出的方法。由于大多数机器学习算法对数据缩放较为敏感，因此我们使用 Scikitlearn 库中的标准缩放器，对提取出的特征向量进行缩放。

4. 结果与讨论

DeCAF 特征和 R-DeCAF 特征的分类准确率汇总于表 1。R-DeCAF 特征由三种线性降维算法获得，即主成分分析、奇异值分解和线性判别分析。首先，我们进一步研究了使用前述三种卷积神经网络的 DeCAF 特征分类性能。如表. 1, 16 19 40× 100× 200× 400× 第三列所示，对于 BreakHis 数据集中的 , , 和 放大倍率数据，VGG- 和 VGG- 的准确率优于 AlexNet，但在 上表现较差。原因可能在于 VGG-16 和 VGG-19 网络的层数及 3×3 滤波器数量；这些网络能够从图像中提取更多细节。然而，对于 400× 数据，VGG-16 和 VGG-19 网络提取的此类细节并非必要，因为其放大倍数更高，图像本身已呈现这些细节。因此，对于高放大倍率数据，AlexNet 是更好的选择。值得一提的是，与预期一致，VGG-16 和 VGG-19 网络在综合所有放大倍数时的准确率更高。此外，40× 数据的准确率最低。这可能是因为，与其他放大倍数相比，40× 数据的感兴趣区域更为复杂且包含更多信息，因此更难以准确分类数据。乳腺癌组织病理学放大倍数图像的复杂程度会影响分类准确率。

其次，我们进一步探讨了基于不同降维算法的 R-DeCAF 特征的分类性能。采用 PCA 和 SVD 生成 R-DeCAF 特征所得到的结果表明，与 DeCAF 特征相比，分类准确率均提高了近 4.3%。例如，使用预训练 AlexNet 处理全部放大倍率数据时，DeCAF 特征、基于 PCA 的 R-DeCAF 特征以及基于 SVD 的 R-DeCAF 特征对应的准确率分别为 85.95%、90.24% 和 90.18%。此外，即使使用 R-DeCAF 特征，对于高放大倍数数据，AlexNet 仍是更好的选择。LDA 得出的结果显示，分类准确率有所下降；由于移除了信息丰富的特征，该方法无法捕捉数据的本质。LDA 准确率较低的主要原因是它忽略了原始数据集的特征数量，且得到的维度（FVS）会小于类别数减一。因此，使用 LDA 技术进行二分类时，我们只有一个特征 [23]。

表 1。 使用不同线性降维方法处理 DeCAF 特征和 R-DeCAF 特征时的分类准确率（%）（最佳结果以粗体标出）。

框架	放大倍数	DeCAF	R-DeCAF（通过线性方法降维）				
		特征向量大小 = 4096	特征向量大小（累积解释方差）	PCA	特征向量大小（累积解释方差）	SVD	LDA
AlexNet（FC6），支持向量机	40×	84.38±1.6	67（0.25）	**88.04±2.1**	67 (0.25)	88.02±2.1	77.62±1.8
	100×	86.16±1.1	45 (0.20)	89.54±1.0	45 (0.20)	**89.66±0.9**	79.87±2.3
	200×	87.69±1.4	97 (0.30)	**90.60±1.4**	135 (0.35)	90.50±1.3	83.87±1.4
	400×	87.91±1.5	23 (0.15)	91.13±1.4	23 (0.15)	**91.15±1.5**	82.58±1.5
	全倍率	85.95±0.8	63 (0.20)	**90.24±0.6**	63 (0.20)	90.18±0.6	71.73±1.1
VGG-16（FC6），支持向量机	40×	86.64±2.2	90 (0.30)	**89.82±1.7**	90 (0.30)	89.60±1.7	82.23±0.9
	100×	89.52±1.2	58 (0.25)	91.01±0.7	206 (0.45)	**91.01±1.2**	84.05±1.8
	200×	88.71±0.9	84 (0.30)	90.78±1.4	56 (0.25)	**90.82±1.1**	82.68±0.9
	400×	85.60±1.9	58 (0.25)	**88.30±1.3**	58 (0.25)	88.21±1.3	81.95±1.9
	全倍率	87.23±0.8	103 (0.25)	**90.61±0.9**	103 (0.25)	90.35±0.8	73.10±1.0
VGG-19 (FC6)，支持向量机	40×	85.09±1.8	78 (0.30)	**87.34±2.0**	51 (0.25)	87.14±1.6	79.62±1.9
	100×	88.06±1.5	118 (0.35)	**90.19±1.3**	118 (0.35)	90.05±1.3	81.10±1.3
	200×	88.51±1.0	114 (0.35)	89.35±1.2	114 (0.35)	**89.60±1.6**	82.80±1.3
	400×	86.73±0.8	59 (0.25)	88.60±0.8	59 (0.25)	**88.76±0.9**	82.03±1.9
	全倍率	86.68±0.5	93（0.25）	**89.43±0.5**	93（0.25）	89.37±0.5	72.03±0.8

为了评估我们所提出的方法，我们计算了 DeCAF 特征和 R-DeCAF 特征分类结果的混淆矩阵。这些矩阵如图5所示，其中展示了预训练 AlexNet 对 BreakHis数据集的 400× data（1820 张图像，其中 588 张为良性，1232 张为恶性）的分类结果。可以看出，通过主成分分析获得的 R-DeCAF 特征的分类结果优于 DeCAF 特征。具体而言，正确预测的良性病例比例从 68% 提高到 82%。这是一个非常令人瞩目的

结果，因为我们所提出的方法能够提高良性类别的准确率，尽管由于 BreakHis 数据集类别不平衡，该类别的数据量相对有限。

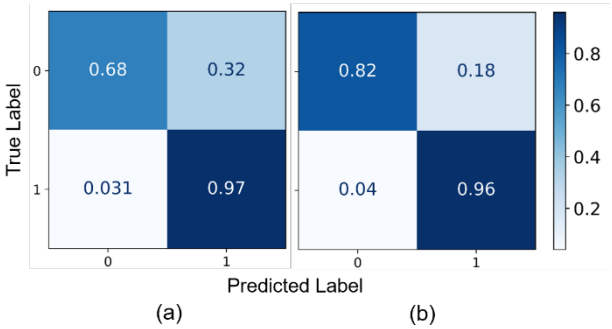


图 5. 使用来自 BreakHis数据集的 400× 数据，展示从预训练 AlexNet 提取的深度特征的分类结果的混淆矩阵。(a) DeCAF 特征；(b) 使用 PCA 算法的 R-DeCAF 特征（FVS=23，CEV=0.15）。0：良性，1：恶性。

此外，我们还使用其他指标验证了 DeCAF 和 R-DeCAF 特征的分类结果，如表 2 所示。我们报告了经 PCA 算法降维后的 R-DeCAF 特征，因为我们观察到 PCA 的准确率优于其他线性降维算法（见表 1）。表 2 显示，与 DeCAF 特征相比，R-DeCAF 特征的精确率和 F1 分数有所提高，但召回率表现不同，在某些情况下，R-DeCAF 特征的召回率有所下降。

由于 BreakHis数据集存在类别不平衡，恶性类别的样本数几乎是良性类别的两倍，而且不同放大倍数下这一比例几乎相同，因此我们采用两种策略，通过复现实验结果来解决这一问题。首先，我们使用加权支持向量机，并采用分层 k 折（k=10）将数据划分为训练集和测试集。其次，我们通过随机抽取恶性样本，使每个放大倍数下恶性样本数与良性样本数相同，从而强制平衡数据，并使用支持向量机。两种策略的结果均表明，表 1 中的分类指标较高；由于数据的不平衡比率并不太高，数据不平衡的影响并不显著。这些研究还证实，平衡数据可能会影响召回率并使其提高，但与表 2 相比，其整体提升幅度较小。

表 2. DeCAF 和 R-DeCAF 特征（经 PCA 降维）的分类准确率、精确率、召回率和 F1 分数（%）。

框架	放大倍数	DeCAF（特征向量大小 = 4096）				R-DeCAF（经主成分分析降维）				
		准确率	精确率	召回率	F1	特征向量大小（累积解释方差）	准确率	精确率	召回率	F1
AlexNet（FC6），支持向量机	40×	84.38±1.6	85.20±2.1	93.42±1.0	89.10±1.2	67 (0.25)	**88.04±2.1**	**89.61±2.7**	**94.73±1.5**	**91.54±1.6**
	100×	86.16±1.1	85.77±1.5	**95.95±0.8**	90.57±0.9	45 (0.20)	**89.54±1.0**	**90.05±1.9**	95.52±1.3	**92.68±0.7**
	200×	87.69±1.4	87.63±1.5	95.58±1.5	91.42±0.9	97 (0.30)	**90.60±1.4**	**91.11±1.8**	**95.65±1.3**	**93.31±1.0**
	400×	87.91±1.5	87.32±2.1	**96.02±1.2**	91.44±1.2	23（0.15）	**91.13±1.4**	**91.31±1.6**	95.96±0.9	**93.57±1.1**
	全倍率	85.95±0.8	86.98±0.9	93.35±0.6	90.05±0.5	63 (0.20)	**90.24±0.6**	**91.24±0.6**	**94.78±0.7**	**92.97±0.5**
VGG-16 (FC6)，支持向量机	40×	86.64±2.2	88.09±2.7	93.17±1.9	90.53±1.7	90（0.30）	**89.82±1.7**	**90.43±1.9**	**95.28±1.8**	**92.77±1.3**
	100×	89.52±1.2	89.21±1.4	**96.53±1.2**	92.72±0.8	58 (0.25)	**91.01±0.7**	**92.15±1.1**	95.11±1.5	**93.59±0.6**
	200×	88.71±0.9	89.23±1.5	95.07±1.6	92.04±0.6	84 (0.30)	**90.78±1.4**	**91.03±1.8**	**96.04±1.1**	**93.46±1.0**
	400×	85.60±1.9	85.43±2.2	95.08±1.4	89.98±1.5	58 (0.25)	**88.30±1.3**	**88.92±1.8**	**95.64±1.4**	**91.67±1.0**
	全倍率	87.23±0.8	87.86±0.9	94.40±0.6	91.01±0.5	103 (0.25)	**90.61±0.9**	**91.30±0.8**	**95.18±0.8**	**93.20±0.7**
VGG-19 (FC6)，支持向量机	40×	85.09±1.8	86.00±2.4	93.51±1.0	89.58±1.3	78 (0.30)	**87.34±2.0**	**87.01±2.2**	**95.87±1.3**	**91.21±1.5**
	100×	88.06±1.5	87.33±2.0	**96.86±0.9**	91.84±1.1	118 (0.35)	**90.19±1.3**	**90.60±1.5**	95.81±1.0	**93.13±1.0**
	200×	88.51±1.0	88.33±1.0	**96.31±0.9**	**92.14±0.6**	114 (0.35)	**89.35±1.2**	**90.10±1.6**	95.29±1.2	92.61±0.8
	400×	86.73±0.8	86.59±1.5	**95.35±1.3**	90.74±0.6	59 (0.25)	**88.60±0.8**	**90.27±0.6**	93.36±1.4	**91.78±0.6**
	全倍率	86.68±0.5	87.39±0.5	94.10±0.7	90.62±0.5	93 (0.25)	**89.43±0.5**	**89.93±0.6**	**95.22±0.5**	**92.50±0.4**

此外，本研究还评估了非线性降维方法的性能，包括第3.3节中介绍的核主成分分析和 t-SNE。表 3 展示了基于 DeCAF 和 R-DeCAF 特征的分类准确率，以及特征向量大小。使用核主成分分析算法时，我们测试了不同数量的特征。然而，基于 R-DeCAF 特征的分类准确率不够高。因此，我们考虑采用与主成分分析方法相近的特征数量。应用 t-SNE 算法时，强烈建议先使用主成分分析方法，再通过 t-SNE 将维数降至 2 或 4 个特征。因此，我们在仅对特征向量应用主成分分析的同一阶段降低了特征向量大小。随后，使用 t-SNE 将特征数量降至 2。显然，非线性降维方法无法有效从 DeCAF 特征中提取信息丰富的特征，导致分类准确率下降。然而，主成分分析等线性方法能够更好地组合深度特征，从而提高分类任务的准确率。非线性降维技术可能对维数灾难较为敏感，这或许是它们在本研究中表现不佳的原因。因此，这些方法无法保证比主成分分析等线性方法取得更好的性能 [23, 27]。此外，我们还可以认为，通过非线性降维方法获得的 R-DeCAF 特征更加复杂，这会导致分类准确率降低。

表3. 使用不同非线性降维方法时 DeCAF 和 R-DeCAF 特征的分类准确率（%）（最佳结果以粗体显示）。

框架	放大倍数	DeCAF 特征向量大小 = 4096	R-DeCAF（通过非线性方法降维）		
			特征向量大小（累积解释方差）	核主成分分析	主成分分析+ t-SNE
AlexNet（FC6），SVM	40×	**84.38±1.6**	67 (0.25)	68.45±2.6	65.94±5.7
	100×	**86.16±1.1**	45 (0.20)	69.33±1.9	68.85±1.5
	200×	**87.69±1.4**	97 (0.30)	68.58±1.8	65.23±4.8
	400×	**87.91±1.5**	23（0.15）	67.47±2.4	67.03±4.7
	全倍率	**85.95±0.8**	63 (0.20)	68.15±1.0	62.96±6.0
VGG-16 (FC6), SVM	40×	**86.64±2.2**	90 (0.30)	68.77±2.8	63.51±11.2
	100×	**89.52±1.2**	58 (0.25)	69.11±1.0	68.94±3.2
	200×	**88.71±0.9**	84 (0.30)	68.67±1.5	63.57±10.9
	400×	**85.60±1.9**	58 (0.25)	68.21±2.5	64.78±6.9
	全倍率	**87.23±0.8**	103 (0.25)	68.46±0.7	66.15±2.7
VGG-19 (FC6), SVM	40×	**85.09±1.8**	78 (0.30)	68.65±2.2	55.93±13.20
	100×	**88.06±1.5**	118 (0.35)	69.45±2.0	63.86±7.8
	200×	**88.51±1.0**	114 (0.35)	69.95±1.4	68.39±3.0
	400×	**86.73±0.8**	59 (0.25)	68.27±1.9	66.73±5.4
	全倍率	**86.68±0.5**	93 (0.25)	68.44±0.9	66.96±2.5

根据我们的分析以及表 1 和表 3 中报告的结果，我们可以得出结论：在不同放大倍数下，使用 PCA 和 SVD 等线性降维算法处理 R-DeCAF 特征，可使 BreakHis数据集的二分类准确率最高提升 4.3%。PCA 算法较低的过拟合概率和噪声抑制能力，以及 SVD 算法在稀疏数据管理方面的优势，[23, 30] 是 R-DeCAF 特征性能提升的原因。这是一项重要发现：深度特征之间存在线性组合，这有助于我们在修改网络时加以考虑，从而提升其性能。

此外，我们将所提出的框架与最先进的研究进行了比较，结果汇总于表 4。我们报告了通过 PCA 算法获得的 R-DeCAF 特征的结果。以往研究从不同的预训练卷积神经网络中提取深度特征，并像我们的方法一样使用 SVM分类器对 BreakHis数据集进行分类。表 4 还列出了特征向量大小，以便进行全面分析和比较。可以看到，与一些方法相比，所提出的方法取得的结果提高了准确率。例如，在 [8] 和 [14], 中，对从预训练卷积神经网络（即 AlexNet、VGG-16 和 VGG-19）提取的深度特征进行分类，所得准确率低于本研究。Kumar等人 [13], 、Gupta等人 [20], 和 Alinsaif 等人 [16] 报告的准确率更高，但其特征向量大小无法与我们的方法相比，几乎低于

120。在 [13], 中，作者对 VGG-16 的全部五个块中的五个外部卷积层应用全局平均池化，并在拼接后得到一个大小为 1472 的特征向量。这种方法与我们不同：我们只提取了一个层（第一个全连接层）的特征。这说明，利用所有层的特征能够改善结果。在 [20] 中报告了更高的准确率，其特征向量大小为 500。不过，特征数量减少时，准确率也会下降。此外，[16] 中的作者只有在特征向量大小为 1300 时，才能保持分类准确率不变。另一方面，由于我们仅分析了三种作为特征提取器的预训练模型，因此无法在所提及的研究 [16, 20] 中检验我们的设想，因为这些研究使用的卷积神经网络特征提取器各不相同。此外，我们认为，所提出的方法能够提升迁移学习的性能。另一项基于微调预训练卷积神经网络的研究报告称，BreakHis数据集的 40× 数据上的分类准确率为 80.80% [5]。将特征向量大小降至 120 以下后，与以往研究相比，所提出的方法取得了理想效果，同时分类准确率提高了 4.3%。

表4. 所提出的方法与以往方法所得分类准确率的比较。

现有方法	CNN	FVS	分类准确率（%）				
			40×	100×	200×	400×	整体 mag.
Bardou 等 [18]	新型 CNN	2000	90.64	89.58	90.23	75.96	-
Deniz 等人 [8]	AlexNet + VGG-16	4096 + 4096	84.87±1.1	89.21±1.4	88.65±2.4	86.75±4.2	-
Gupta 等人 [20]	ResNet	500	97.07±1.2	96.10±1.0	94.69±1.2	90.85±2.1	-
Kumar 等人 [13]	VGG-16	1472	94.11±1.8	95.12±1.1	97.01±1.1	93.40±1.0	-
Saxena 等人 [14]	AlexNet	1526	84.06	87.54	89.40	85.16	-
	VGG-16	3072	86.36	87.77	86.80	84.35	-
	VGG-19	3072	86.64	88.17	85.84	81.67	-
Alinsaif 等. [16]	DenseNet	1300	-	-	-	-	97.96±0.6
所提出的	AlexNet	23-97	88.04±2.1	89.54±1.0	90.60±1.4	91.13±1.4	90.24±0.6
	VGG-16	58-103	89.82±1.7	91.01±0.7	90.78±1.4	88.30±1.3	90.61±0.9
	VGG-19	59-118	87.34±2.0	90.19±1.3	89.35±1.2	88.60±0.8	89.43±0.5

5. 结论

本研究提出了用于基于组织病理图像进行乳腺癌检测的 R-DeCAF 特征，并将其与 DeCAF 特征进行了比较。为提取 DeCAF 特征，我们采用了三种不同的预训练卷积神经网络作为无监督特征提取器。我们从卷积神经网络的第一个全连接层提取了特征向量，其 FVS 为 4096。结果表明，在分类任务中，保留从预训练 AlexNet、VGG-16 和 VGG-19 提取的全部 DeCAF 特征并不奏效。因此，我们评估了多种针对 DeCAF 特征的降维方法，以提取信息丰富的特征向量，同时缩短计算时间。根据我们的分析，在新空间中保留约 15% 至 35% 的特征累积解释方差，且 FVS 小于 120，已足够，并且在最佳情况下可将准确率显著提高 4.3%。评估结果表明，线性降维算法能够有效结合深度特征，并提高分类任务的准确率；然而，非线性方法则未能奏效。这是一项重要发现：深度特征之间存在线性组合关系，这有助于我们在改进网络时加以利用，从而提升性能。此外，在多种线性降维方法中，主成分分析表现最佳。使用预训练 AlexNet 作为特征提取器时， 400× data 的最佳结果为 91.13±1.4%。值得注意的是，所提出的模型无需数据增强或专门的数据预处理，可视为一种全自动癌症诊断模型。此外，BreakHis数据集的放大倍率会影响分类准确率，因为它取决于组织病理图像的复杂程度。未来的工作可基于 PCA 算法改进深度卷积神经网络模型，以降低特征复杂度，并利用更可靠、信息更丰富的特征提高分类准确率，从而有望打破这一瓶颈。此外，还将进一步研究其他预训练卷积神经网络模型，用于提取深度特征，并应用所提出的方法提升性能。

，这将作为未来工作的研究方向。

参考文献

[1] T. Araújo、G. Aresta、E. Castro、J. Rouco、P. Aguiar、C. Eloy、A. Polónia 和 A. Campilho，《使用卷积神经网络对乳腺癌组织学图像进行分类》，*PLoS One*，第12卷，第6期，文章编号e0177544，2017年。[2] W. H. Organization，《世界卫生组织关于乳腺*X*线筛查的立场文件》。世界卫生组织，2014年。[3] P. Boyle 和 B. Levin，《*2008*年世界癌症报告》。IARC出版社，国际癌症研究机构，2008年。[4] J. Arevalo、A. Cruz-Roa 和 F. A. GONZÁLEZ O，《用于自动分析的组织病理图像表示：最新综述》，*Revista Med*，第22卷，第2期，第79–91页，2014年。[5] S. Singh 和 R. Kumar，《利用深度Inception模块和残差模块从组织病理图像中检测乳腺癌》，*Multimedia Tools and Applications*，第81卷，第4期，第5849–5865页，2022年。[6] F. A. Spanhol、L. S. Oliveira、P. R. Cavalin、C. Petitjean 和 L. Heutte，《用于乳腺癌组织病理图像分类的深度特征》，载于*2017*年*IEEE*系统、人类与控制论国际会议（*SMC*），2017年：IEEE，第1868–1873页。[7] R. Mehra，《乳腺癌组织学图像分类：从头训练还是迁移学习？》，*ICT Express*，第4卷，第4期，第247–254页，2018年。[8] E. Deniz、A. Şengür、Z. Kadiroğlu、Y. Guo、V. Bajaj 和 Ü. Budak，《基于迁移学习的乳腺癌检测组织病理图像分类》，*Health Information Science and Systems*，第6卷，第1期，第1–7页，2018年。[9] G. Zhong、S. Yan、K. Huang、Y. Cai 和 J. Dong，《通过缩减和拉伸深度卷积激活特征实现准确的图像分类》，*Cognitive Computation*，第10卷，第1期，第179–186页，2018年。[10] P. Filipczuk、T. Fevens、A. Krzyżak 和 R. Monczak，《基于细针活检细胞学图像分析的计算机辅助乳腺癌诊断》，*IEEE*医学影像汇刊，第32卷，第12期，第2169–2178页，2013年。[11] S. Sharma 和 R. Mehra，《用于乳腺癌组织病理图像多分类的传统机器学习与深度学习方法——比较性分析》，*Journal of Digital Imaging*，第33卷，第3期，第632–654页，2020年。[12] T. J. Alhindi、S. Kalra、K. H. Ng、A. Afrin 和 H. R. Tizhoosh，《比较LBP、方向梯度直方图和深度特征在组织病理图像分类中的表现》，载于*2018*年国际神经网络联合会议（*IJCNN*），2018年：IEEE，第1–7页。[13] A. Kumar、S. K. Singh、S. Saxena、K. Lakshmanan、A. K. Sangaiah、H. Chauhan、S. Shrivastava 和 R. K. Singh，《用于犬乳腺肿瘤和人类乳腺癌组织病理图像分类的深度特征学习》，*Information Sciences*，第508卷，第405–421页，2020年。[14] S. Saxena、S. Shukla 和 M. Gyanchandani，《将预训练卷积神经网络作为特征提取器用于基于组织病理学的乳腺癌诊断》，*International Journal of Imaging Systems and Technology*，第30卷，第3期，第577–591页，2020年。[15] P. Yamlome、A. D. Akwaboah、A. Marz 和 M. Deo，《基于卷积神经网络的乳腺癌组织病理图像分类》，载于*2020*年第*42*届*IEEE*医学与生物学工程学会年度国际会议（*EMBC*），2020年：IEEE，第1144–1147页。[16] S. Alinsaif 和 J. Lang，《使用深度特征和迁移学习进行组织学图像分类》，载于*2020*年第*17*届计算机与机器人视觉会议（*CRV*），2020年：IEEE，第101–108页。[17] S. Boumaraf、X. Liu、Y. Wan、Z. Zheng、C. Ferkous、X. Ma、Z. Li 和 D. Bardou，《传统机器学习与深度学习在依赖放大倍率的乳腺癌组织病理图像分类中的比较：结合可视化解释的比较研究》，*Diagnostics*，第11卷，第3期，文章编号528，2021年。[18] D. Bardou、K. Zhang 和 S. M. Ahmad，《基于组织学图像并使用卷积神经网络的乳腺癌分类》，*IEEE Access*，第6卷，第24680–24693页，2018年。[19] M. Z. Alom、C. Yakopcic、M. Nasrin、T. M. Taha 和 V. K. Asari，《利用Inception循环残差卷积神经网络从组织病理图像中进行乳腺癌分类》，*Journal of Digital Imaging*，第32卷，第4期，第605–617页，2019年。

[20] V. Gupta 和 A. Bhavsar，《用于乳腺癌组织病理图像分类的部分独立框架》，载于《IEEE/CVF计算机视觉与模式识别会议研讨会论文集》，2019年，第0–0页。[21] 《乳腺癌组织病理数据库（BreakHis）》。https://web.inf.ufpr.br/vri/databases/breast- cancer-histopathological-database-breakhis/（访问日期: [22] R. F. Mansour，《基于深度学习的自动计算机辅助糖尿病视网膜病变诊断系统》，《生物医学工程快报》，第8卷，第1期，第41–57页，2018年。[23] F. Anowar、S. Sadaoui 和 B. Selim，《降维算法（主成分分析、核主成分分析、线性判别分析、多维尺度分析、奇异值分解、局部线性嵌入、等距映射、拉普拉斯特征映射、独立成分分析、t-SNE）的概念与实证比较》，《计算机科学评论》，第40卷，文章编号100378，2021年。[24] K. P. Murphy，《机器学习：概率视角》。麻省理工学院出版社，2012年。[25] A. Krizhevsky、I. Sutskever 和 G. E. Hinton，《使用深度卷积神经网络进行ImageNet分类》，《神经信息处理系统进展》，第25卷，2012年。[26] K. Simonyan 和 A. Zisserman，《用于大规模图像识别的超深度卷积网络》，arXiv预印本arXiv:1409.1556，2014年。[27] L. Van Der Maaten、E. Postma 和 J. Van den Herik，《降维：一项比较研究》，《机器学习研究杂志》，第10卷，第66–71期，文章第13号，2009年。[28] A. Tharwat、T. Gaber、A. Ibrahim 和 A. E. Hassanien，《线性判别分析：详细教程》，《人工智能通信》，第30卷，第2期，第169–190页，2017年。[29] L. Van der Maaten 和 G. Hinton，《使用t-分布随机邻域嵌入可视化数据》，《机器学习研究杂志》，第9卷，第11期，2008年。[30] S. Karamizadeh、S. M. Abdullah、A. A. Manaf、M. Zamani 和 A. Hooman，《主成分分析概述》，《信号与信息处理杂志》，第4卷，2020年。