

用逐步微调深度神经网络进行胃癌病理图像分类

摘要：用卷积神经网络（CNNs）的深度学习对于图像分类任务的一个重要工具。由于其有出色的鲁棒性和泛化性，也有望为病理图像的高级计算机辅助诊断（CAD）发挥关键作用。然而，由于对病理医生专业观察的高成本标注，目前缺乏用于训练深度神经网络的良好标注的病理图像数据已成为一个主要问题。面对这个问题，迁移学习技术通常用来增强深度神经网络的能力。为了进一步提高目前最先进的深度神经网络的性能，缓解良好标注数据的不足，本文提出了一种新的基于逐步微调的胃癌病理图像分类深度学习方案，并建立了一种新的目标相关的中间数据集。我们提出的方案能够使深度神经网络模仿病理医生的感知方式，提前获得病理相关知识，但在数据标注上的额外成本非常少。这实验是在几个当前比较好的深度神经网络上，对标注良好的胃癌病理数据和目标相关的中间数据两个数据集之间进行的。结果一致地证明了我们提出的方案在提高分类性能方面的可行性和优越性。

1. 介绍

众所周知，癌症被认为是人类健康最大的威胁之一。根据国际癌症研究机构（IARC）统计，在 2012 年，全球将近有 1410 万新癌症患例和 820 万人死亡。到 2035 年，预计将增加到 2400 万，并且死亡率在不断的提升。在诊断方法上，虽然先进的图像诊断设备（CT, MRI, US）正在快速地发展，但对于许多癌症来说，病理诊断仍然被认为是评估癌症有无、类型、恶性程度的黄金标准。然而，病理医生的缺乏对病理学的发展有很大的限制，引发了很多社会问题。在美国，病理医生的短缺已经变成越来越值得关注的事。在日本，病理医生的数量还不到美国的三分之一（每 19000 人有一个）。这个情况在中国更为严峻。据报道，中国大约每 74000 人才有一个病理医生。因此，这种严重的短缺现在给病理医生带来了巨大的工作压力，并可能造成诊断上的错误和疏忽。对比 2005 年，2012 年日本全部病理诊断病例数和术中病理诊断病例数分别上升到 1.7 倍和 3 倍。类似数字诊断的新技术广泛迅速的使用，由于虚拟显微镜使操作更简单易用，诊断变得更为便宜。然而，由于诊断的正确性和减轻病理医生的工作量仍然面临挑战，基于先进图像分类技术的进一步帮助有望发挥关键作用，促进更先进的计算机辅助病理诊断。

早期常用的病理图像分类方法包括特定的组织学特征或广义纹理特征。特定的组织学相关特征，如细胞核的面积和核质比（N/C），能够从未知图像中精确地计算出来。这些特征与预定义的标准进行比较，判断目标图像是良性肿瘤还是恶性肿瘤。不幸的是，这样的过程通常会遇到一个大问题，那就是很难对形态学特征作出充分的定义，因为癌细胞通常缺乏对正常分裂的控制。因此，细胞形状提取失败可能成为分类失败的直接原因。相比之下，使用广义纹理图像特征的方案似乎对各种癌变现象保持了更强的鲁棒性。作为一种常用的方法，以往的许多研究都充分利用了广义纹理特征，证明了它们对不同任务的处理能力。灰度共生矩阵（GLCM）是纹理特征的一种。如，Esgiar 等人利用 GLCM 从结肠活检中获得与对比度、熵、角二矩阵、差异性、相关性对应的纹理特征，利用线性判别分析（LDA）和 k 邻近算法（KNN）实现正常结肠黏膜和癌变结肠黏膜的分类。Diamond 等人可能使用 Haralick 特征（一种由 GLCM 发展而来的纹理特征）来识别前列腺病理图像中的组织异常。另一个有力的对手是本地二进制模式（LBPs）。在 Masood

和 Rajpoot 的研究中，提出了一种由 LBP 和支持向量机 (svm) 组成的结肠病理图像处理方案，并证明了该方案的有效性。在另一项工作中，Sertel 等人开发了神经母细胞瘤 H&E 染色全切片图像的分类方法，是用来与上述研究相似的共生统计和局部二进制模式。Kather 等人最近的一篇报道对结直肠癌组织图像的纹理分析进行了较为全面的研究。除了 LBP 和 GLCM，还涉及到低阶和高阶直方图特征、Gabor 过滤器和类似感知的特征。在我们早期的研究中，还发现了另一种纹理特征，称为高阶局部自相关 (HLAC)，与基于主成分分析 (PCA) 的子空间方法等线性统计模型相结合，能够显示胃癌病理图像的异常程度。除了单纯的良恶性分类外，病理图像领域也提出了一些具有纹理特征的分类方法，解决了类似的分类相关任务，如腺体分割和等级估计等。

虽然所有这些基于纹理特征的方法都显示出了良好的可行性，但是在研究和实际应用之间仍然存在一些难以解决的问题。一个特殊的例子是确定手工制作的几何特征是否适合某些任务是相当困难的。同时，图像间 H&E 染色不均匀会对分类性能产生不利影响，使任务更具挑战性。自 2012 年 ImageNet 大型视觉识别大赛 (ILSVRC) 上使用深度学习的团队取得压倒性胜利以来，许多图像识别技术已经被使用卷积神经网络 (CNNs) 的深度学习所取代。由于更多的领域不可知方法结合了特征发现和实现来最大限度地区分感兴趣的类，深度学习对各种图像显示出前所未有的适应性。因此，人们对深度学习寄予厚望，希望它能在病理图像等医学图像分析领域发挥巨大的作用。特别是在病理图像领域，除了渴望更精确的分类和分割，深度学习也被启发用于新的扩展应用，如染色归一化、评估肿瘤扩散度、医学图像与诊断报告之间的综合多模态映射。

与自然图像的计算机视觉相比，医学图像领域训练数据的缺乏以及准确的标注成为当前的首要问题。由于医生知识和协作的必要性，获取过程通常既要耗费昂贵的财力，又要大量的工作量。对于这个问题，许多研究已经证明，使用深度神经网络微调技术迁移学习在某种程度上可以提高性能和减少训练数据的不足影响，迁移一个常用的神经网络预训练大规模图像数据集（比如 ImageNet），指定一个对应于更复杂的目标任务。然而，由于预训练的图像数据集具有固定的内容类别和图像大小，对于特定的图像分类领域中遇到的问题，数据集和深度神经网络的系数效率一般都是次优的。在许多情况下，从预训练的任务中获得的知识与不同的指定目标领域之间仍然存在差距。

因此，本文旨在进一步减轻对胃癌病理图像良好的标注不足和用理性的方法提升性能，我们提出了一种新颖的方案，对 CNNs 采用两级微调方法，介绍了一种新型目标相关的中间数据集（以下称“中级”数据集）。除了“低级”的大规模预训练数据集数量庞大但目标特异性较低，以及“高级”注释良好的病理数据集与目标任务直接相关外，本文提出的“中级”数据集是基于病理图像领域中的组织和细胞信息生成的。利用“中级”数据集，我们的方案能够使深层神经网络模仿病理学家的感知方式，提前获取病理学相关知识，但在数据标注方面的额外成本非常有限。接下来，我们将给出逐步微调方案和数据集的详细资料。之后，我们会进行实验证明我们提出的方案。

2. 材料和方法

这部分将展开几个子主题。首先，我们将简要介绍卷积神经网络 (CNNs) 和

本文采用的体系结构。其次，我们将给出一些关于逐步微调方案的细节（图1），以阐明它如何有助于提高 CNN 的性能，并减少好标注的病理训练数据的不足问题。

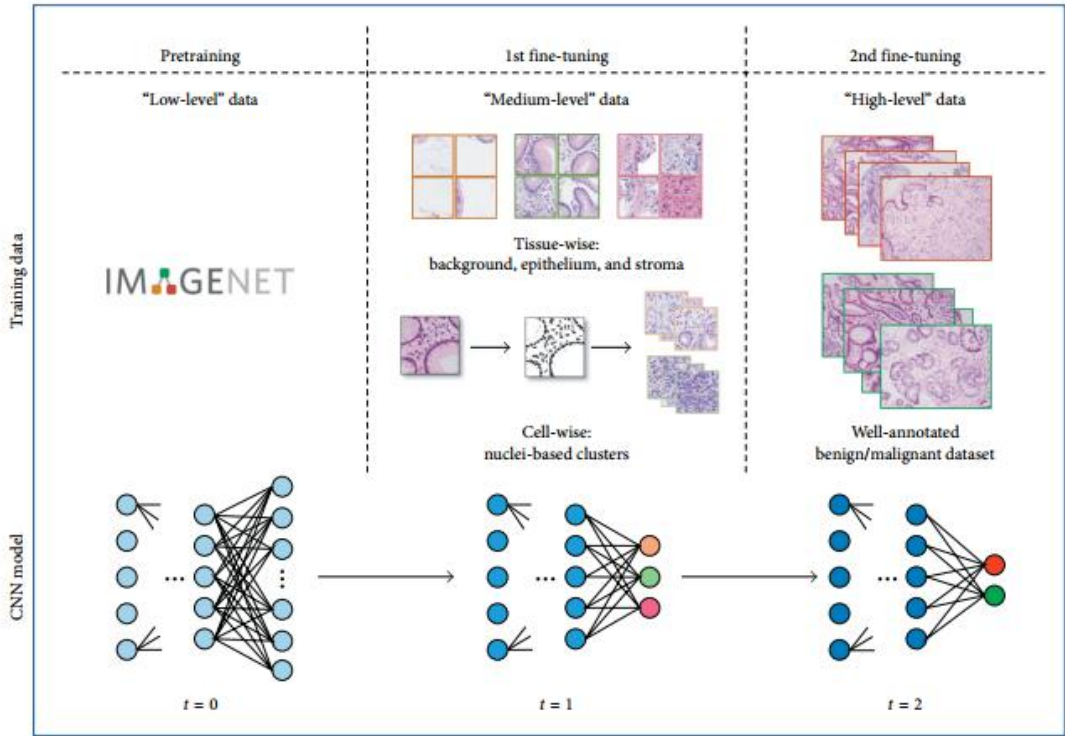


图 1：逐步微调结构。除了“低级”预训练数据集和好标注的“高级”数据集，还包括组织和细胞数据的“中级”数据集。在与训练步骤 $t=0$ ， $t=1$ 和 $t=2$ 对应的 CNN 模型中，颜色较深的蓝色节点表示病理图像分类任务的更具体（更深）表示。

第三，我们将展示一个可行的组织层面的“中级”数据集，包括背景、上皮和基质的类别，这些数据集可以基于病理知识进行半专业标注，只花很少的工作量。最后，为了使标注成本最小化，并在极度不足的标注劳动力提供一种可选的解决方案，我们进一步受到启发，提出了一种解决方案，通过细胞核测量和无监督聚类实现了细胞“中级”数据集的全自动产生。

2.1 卷积神经网络（CNN）结构。在图像识别和计算机视觉领域研究的各种方法中，卷积神经网络（CNN）是目前最成功的。CNN 的原型可以在基于生物大脑视觉皮层的神经生理学发现而设计的认知子中找到。它是一种神经网络，交替排列与细胞对应的卷积层进行特征提取，池化层对应于允许分层位置偏差功能的细胞。直观地说，它可以理解为一个网络，在不同的尺度上，将相邻特征一点点的共现，有选择地向上层提供有效的识别信息。实际上，这种信息的细化通常是通过减少损失函数来实现的：

$$L(W) = \frac{1}{N} \sum_{i=1}^N l(y(x_i; W), y_i). \quad (1)$$

式（1）中， $L(W)$ 为 N 个训练样本数据集的总损失（当前预测值与实际值的差值），权值 W ， y_i 为训练数据 x_i 的标签。 $Y(x_i; w)$ 为 x_i 的预测标签， l 为损失函数。实际上，为了评估所提出的逐步微调方案的有效性和泛化，在本文中，我们采用了几种最先进的 CNN 架构，包括 VGG-16、AlexNet 和 GoogleNet（以下

inception V3)。

2.2 让 CNN 通过逐步微调学习病理学。微调方法在图像分类领域得到了广泛的应用，它通常表示对目标任务的数据集预先训练好的 CNN 进行再训练。由于训练 CNN 对初始参数的依赖性很强，为了防止学习过拟合，尽可能获得合适的初始化参数是非常重要的。一般来说，CNN 的早期图层负责获取相对通用的图像特征，这些特征被认为与传统的纹理特征类似，适用于许多相关的任务，而后期的图层则涉及到与目标任务相对应的更具体的信息。因此，如果用于训练前的数据集和目标数据集之间的分布差异足够相关，在许多情况下，可以对部分或所有层进行微调，以获得比从零开始训练更理想的结果。然而，在其他情况下，如果目标任务与训练前数据集分布不同，则初始化和微调的有效性可能受到很大的限制。这正是病理图像分类中的问题。从人类的感知来看，病理图像通常比自然图像（包括训练前的数据集中）具有更复杂的外观，因为 H&E 染色由于其颜色的均匀性，很难直观地分辨良恶性图像之间的差异。从这个角度来看，我们认为有必要做一些实质性的努力来填补预训练数据集和标注良好的病理数据集任务之间的“空白”。

因此，本文提出了一种新的概念，如果能够使深度神经网络学会以病理医生的方式理解病理图像。对于病理医生的感知方式和学习进度，病理医生在对病理图像做出良恶性判断前，应事先了解基本组织结构的差异。之后，病理医生可以更专注于详细的形态学特征，包括但不限于细胞的扩散状态和密度，细胞核畸变程度，细胞核大小，核质比。综上所述，病理医生最终可以做出良恶性判断。为了实现这一概念，我们建立了一个逐步微调的方案来传授深度神经网络上的病理知识，并引入了一种新型的目标相关的“中级”数据集。而不是直接使用深度神经网络学习如何判断良恶性，使其从“中级”数据集中获得基本的病理知识，可能有助于完成难度较高的任务。因此，基于固有的框架，将“中级”数据集的训练放在逐步微调方案的中间。如图 1 所示，在训练进度 $t=0$ 开始时，我们有一个预训练网络作为初始化。 $t=1$ 是微调我们提出的“中级”数据集网络。在这一步之后，深度神经网络被认为具有更多的病理相关权重。最后，使用标注良好的良/恶性图像进行二次微调（ $t=2$ ）。在图的下半部分，与训练步骤 $t=0$ 、 $t=1$ 、 $t=2$ 相对应，CNN 模型中颜色较深的蓝色节点表示病理图像分类任务更具体（更深）的表示。当输出类的数量发生变化时，网络结构需要相应地进行调整。

2.3 建立一个低成本的“中级”数据集。对于“中级”数据集，需要满足一个先决条件，即能够比病理医生花低得多的成本标注数据集，同时还要具备与病理相关的基础知识。因此，我们首先关注寻找一种可以由“非专业”人员制作的病理图像数据集。本文提出了一种可行的基质上皮数据集，该数据集仅在病理医生的指导下即可自行构建。上皮细胞和基质是两种组织类型，可以在每个器官中发现。在胃癌病理学领域，上皮组织排列在胃黏膜的外表面，基质组织位于上皮之下。由于上皮细胞与基质之间的肿瘤转移被认为是癌症进展过程中不可避免的，所以这两种类型的组织通常都是在活检检查中提取出来，并在病理图像中显示出来。虽然很难确定这两种组织是否适合非专业人员，但实际上，我们发现描述它们之间视觉上差异是一项相当简单的工作。如图 2 所示，虚线包围的上半部分为上皮组织，下半部分为基质组织。

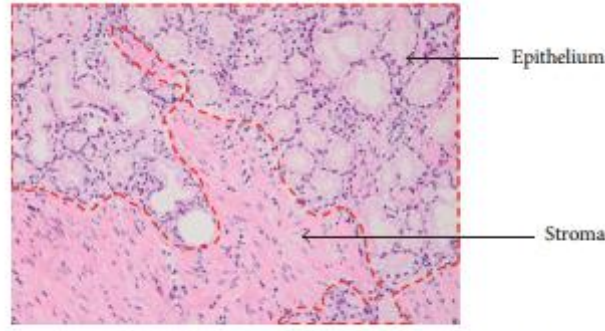


FIGURE 2: Stroma tissue and epithelium tissue appear in gastric pathology image.

胃癌病理图像中基质组织和上皮组织

由于其各自的功能，上皮组织明显表现为有组织的排列，而基质组织则表现为分散无序。这种明显的差异对我们“非专业”工作非常有益。如果不需要病理医生高成本的标注，这些上皮细胞和基质图像能够在我们假设的基础上将病理学知识传授给深度神经网络，那将是令人鼓舞的。实际上，我们已经成功地收集了大量的图像，包括上皮细胞和基质细胞，用基质区域和上皮细胞进行分割，并手工将其分为三类（上皮细胞、基质和背景）。

2.4 自动生成“中级”数据集。在本节中，对比于数据标注工作人员严重不足的情况，本文进一步提出了另一种全自动的“中级”数据集生成方法。该方法首先应用一系列图像处理技术，提取低维特征向量索引测量细胞核的形态。最后，我们应用无监督聚类，形成“中级”数据集。

为了实现对细胞核的测量，需要生成一个二值核图像代替原始核图像。因此，第一步是对 H&E（苏木精和伊红）染色病理图像进行去卷积，将 H&E 染色病理图像分离为 H 和 E 两种成分，因为 H 成分中往往含有丰富的细胞核。该论文采用 Ruifrok 和 Johnston 提出的算法。根据他们的工作，虽然在每个 RGB 通道相对强度取决于非线性的染色剂浓度，但光学密度 $OD_c = -\log_{10}(I_c/I_0, C) = A \cdot b_c$ (c 是任意 RGB 通道) 和吸收系数为 b_c 的染色 A 的浓度是线性的，因此可用来多个染色的分离。实际上， I_c 是 RGB 通道之一的强度， I_0 和 c 值为 1。去卷积矩阵 D 记为 (2)，H, E 和 DAB 的组合可被计算为 $C = D[y]$ ，其中 y 表示 RGB 空间的光密度向量。

$$D = \begin{bmatrix} 1.88 & -0.07 & -0.60 \\ -1.02 & 1.13 & -0.48 \\ -0.55 & -0.13 & 1.57 \end{bmatrix}. \quad (2)$$

接下来，我们使用 Phansalkar 等人提出的自动局部阈值算法对 H（苏木精）通道进行二值化。在该方法中，阈值 T 计算为

$$T = \mu^* \left(1 + p^* e^{-q^* \mu} + k^* \left(\left(\frac{\sigma}{r} \right) - 1 \right) \right). \quad (3)$$

在 (3) 中， μ 和 σ 是本地的平均值和标准偏差。K=0.25, r=0.5, p=2, q=10, 是推荐的常数值。图 3 显示了二值化的结果。

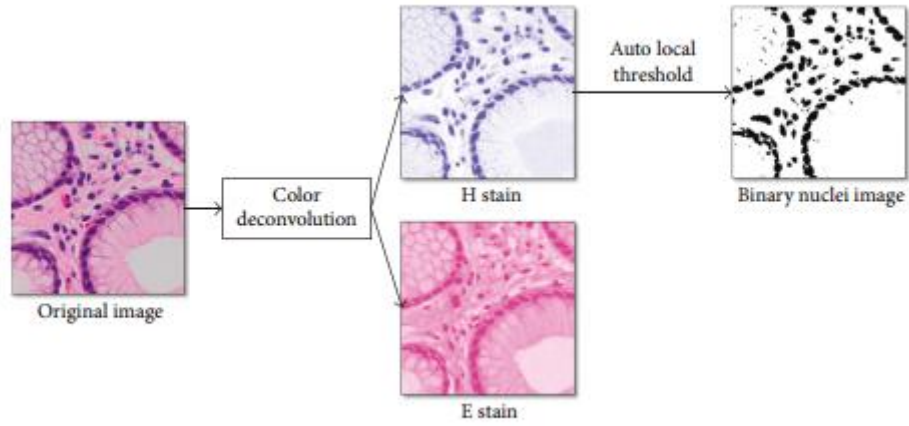


FIGURE 3: Image processing for making binary nuclei images.

然后，我们提出的方法采用分水岭算法分离连结或部分重叠的细胞核，然后采用 Suzuki 和 Abe 提出的轮廓检测方法，从每张病理图像中获取一个总面积和核数的二维向量。毕竟，使用无监督 k 均值聚类将图像分成两类。聚类图像在第一次微调时直接作为“中级”训练数据使用。与组织数据集相比，自动生成的数据集更侧重细胞特征。

3. 结果与讨论

3.1 实验过程。为了评估我们提出的逐步微调方案和低成本的“中级”数据集的有效性，实验按以下步骤进行。(1) 首先讨论了基于不同 CNN 架构的两级微调方案在使用组织数据集时的性能。(2) 然后我们用自动生成的细胞数据集替换组织数据集，并以与 (1) 相同的方式验证方案。在 (1) 和 (2) 中，我们将进一步讨论我们所提出的两级方案对于不同大小良好标注的数据集是如何进行的，以此支持我们的概念。

3.2 数据集。这篇论文使用了“低级”、“中级”和“高级”三种数据类型（图 1）分别用于初始化、一级微调和两级微调。在实践中，ImageNet 数据通常用于初始化 CNN 模型，其中包含 1000 个不同类别的大约 120 万张图像。至于胃癌病理学图像中，所有使用的数据集如表 1 所示。

TABLE 1: Datasets used in experiments.

Data type	Category	Training	Validation	Test
Tissue-wise data	Background	15,000	1,000	—
	Epithelium	15,000	1,000	—
	Stroma	15,000	1,000	—
Cell-wise (nuclei) data	Cluster 1	6,905	767	—
	Cluster 2	5,811	646	—
Well-augmented data	Small	540 + 540	1620 + 1620	2,700 + 2,700
	Large	5400 + 5400		

通过绘制上皮细胞核基质的图谱，我们成功地收集了 48000 个组织块 (256x256)，这些块被分为“背景”、“上皮”和“基质”类别（图 4）。每个类别中使用 15000 个块作为训练数据，其余 1000 个块用于验证。从该数据集中，我们进一步得到了符合细胞核测量的细胞数据集。该细胞数据集由 7672 个块和 6457 个块组成，分为两个簇（图 5），剔除了细胞核过少或由于染色不均匀不易提取的块。

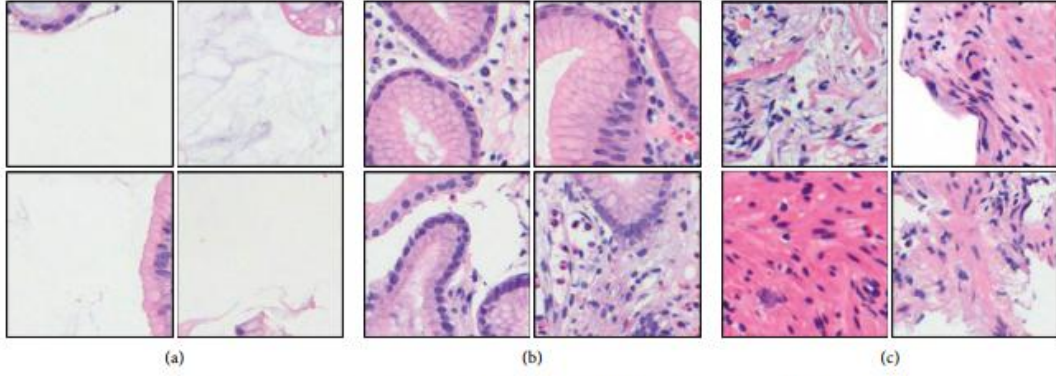


FIGURE 4: “Nonprofessionally” annotated tissue-wise datasets. (a) Background. (b) Epithelium. (c) Stroma.

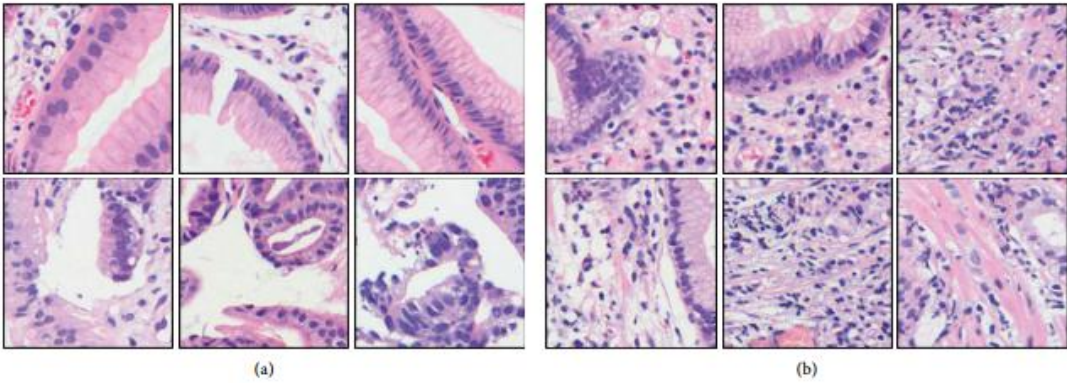


FIGURE 5: Full-automatically generated cell-wise (nuclei) datasets. (a) Cluster 1. (b) Cluster 2.

实验中，每个簇 90% 的块用来训练，其余的用于验证。如图 5 所示，根据粗略的观察，左边的簇似乎有更大更少的核，而右边的簇似乎相反，有更多更小的核。对于良好标注的“高级”数据集，为了评估提出的两级方案的有效性和泛化性，我们准备了两种不同大小的标注良好的数据集。一个是包含 540 个良性和 540 个恶性块的小训练数据集。另一个是包含该小数据集，有 5400 个良性和 5400 个恶性块的非增广数据集。除了之前的数据集，我们还使用了一个包括 1620 个良和 1620 恶性块验证数据集，用来选择最好的配置，以及一个包括 2700 个良和 2700 个恶性块测试数据集，用来最终评估每个情况下的性能。值得注意的是，“中级”数据集和“高级”数据集之间没有重叠，训练，验证，测试数据集间也没有重叠。

4. 实验结果

4.1 使用组织数据集逐步微调。在本部分中，我们将具体讨论我们提出的逐步微调方案，组织数据用于第一级微调，良好标注的数据集用于第二级微调。不同大小和不同深度神经网络结构对病理图像数据集进行实验，评估性能。在本文中，我们同时采用 AUC、ACC、精度和 recall 作为评价标准。

值得注意的是，在表 2 中，在所有方案中，我们提出的使用组织数据进行两级微调的方案都有提升。具体来说，在小数据组的结果中，当我们分别用 CNN 架构 VGG-16、AlexNet 和 Inception V3 时，AUC 值分别提高了 0.035、0.016 和

0.039。在大数据组中，对应的 AUC 值分别提高 0.027、0.053 和 0.053。虽然使用更小的训练数据的性能可以得到更大的提高，但是根据 AUC 值，我们发现我们提出的方案实际上为大型标注良好的数据集组带来了更多的好处。与此同时，如果我们关注 ACC 值，就能发现，当在大数据集中使用 Inception V3 方案时，精度显著地提高了，从 0.779 提高到 0.862。此外，医学图像分类中常用的精度和 recall 指标也有类似于 AUC 和 ACC 的趋势。如图 6 所示，三个 CNN 架构结合两个数据集产生了六个 ROC 图。红色曲线表示使用组织“中级”数据集的两级微调方案，绿色曲线表示传统的一级方案。从图中可以看出，无论是假阳性率轴还是真阳性率轴，我们提出的方案都具有压倒性的区域。结果表明，我们出的方案具有较强的适应性，对良好标注的数据量依赖较少。此外，由逐步训练的网络导出的一组经过过滤的响应图像如图 7 所示。考虑到实际和直观的工具，我们研究了 VGG-16 模型，并从所有卷积块中得到了第一个卷积层的一些结果，如图所示。

TABLE 2: Performances of the proposed two-stage fine-tuning using tissue-wise data.

Data size	Scheme	CNN architecture											
		VGG-16				AlexNet				GoogLeNet (Inception V3)			
		AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall
Small	One stage	0.879	0.793	0.863	0.695	0.828	0.723	0.74	0.72	0.838	0.753	0.75	0.75
	Two stage (tissue)	0.914	0.829	0.865	0.781	0.844	0.761	0.76	0.76	0.877	0.772	0.79	0.77
Large	One stage	0.936	0.836	0.957	0.703	0.867	0.794	0.80	0.79	0.881	0.779	0.79	0.78
	Two stage (tissue)	0.963	0.881	0.869	0.898	0.920	0.837	0.84	0.84	0.934	0.862	0.86	0.86

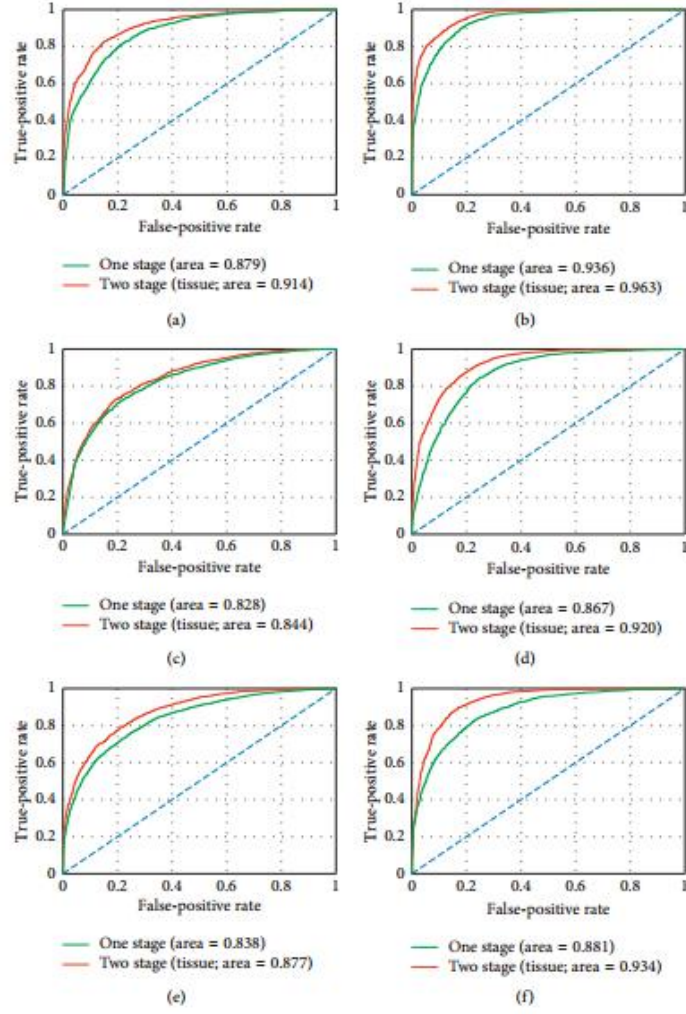


FIGURE 6: Performances of the proposed two-stage fine-tuning using tissue-wise data presented by ROC. (a) VGG-16 + small data. (b) VGG-16 + big data. (c) AlexNet + small data. (d) AlexNet + big data. (e) Inception V3 + small data. (f) Inception V3 + big data.

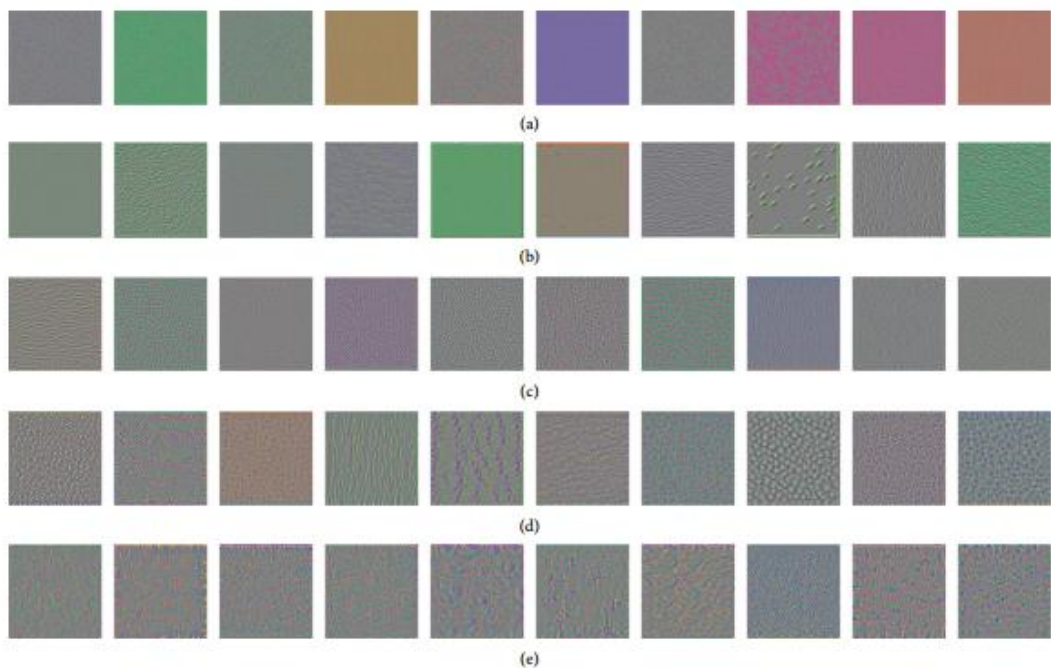


FIGURE 7: A set of filtered response images outputted by the stepwise trained VGG-16 model. (a) Block 1, Conv1. (b) Block 2, Conv1. (c) Block 3, Conv1. (d) Block 4, Conv1. (e) Block 5, Conv1.

根据常识，训练数据越大，AUC 绝对值越大。然而，如果我们对数据大小横向对比，可以看出，当我们对小数据集使用两级微调方案，性能已经接近于当只使用一级微调对十倍于小数据集的大数据集的水平。观察“Small-Two stage(issue)”和“Large-One stage(issue)”不难发现，两个 AUC 之间的差异不大于 0.023 (AlexNet)，而 ACC 的最大差异值只有 0.033 (AlexNet)。这些结果充分证明，组织信息的引入确实有助于加深神经网络对病理图像的“理解”。因此，可以推断，当好的标注数据集有限时，但“非专业”标注在某种程度上可行时，我们提出的两级微调方案可以用来增强性能。

4.2 使用细胞数据集进行逐步微调。在接下来的步骤中，我们将遵循与上面相同的步骤，但是自动聚类方法生成的细胞数据代替组织数据。为了对各个方案进行全面比较，同样在表 3 中，我们列出来两种方案的结果，包括两种大小的标注良好的数据集和三种类型的 CNN 架构。从表 3 中可以看出，使用细胞数据集的逐步微调方案同样适用于所有结果。根据 AUC 值，当使用小的标注良好的数据集时，通过使用细胞核数据集的微调，水平方向上，性能分别提升了 0.023 (VGG-16)、0.024 (AlexNet) 和 0.034 (Inception V3)。在大数据组中，三个网络的 AUC 值分别提高了 0.029 (VGG-16)、0.048 (AlexNet) 和 0.052 (Inception V3)。与 AUC 值一样，无论采用哪种 CNN 架构和良好标注数据集，在我们对细胞核数据集进行第一级微调后，ACC 值都得到了很大的提高。在索引精度和召回率方面，在每一对小数据组中，我们提出的方案都提高了精度和召回率。然而，在大数据组中显示精度降低，相反，作为偏移量的召回率显著增加。根据定义，较大的召回率表示未检测到的异常更少（假阴性），而较大的精度表示过检测更少（假阳性）。实际上，由于未检测到异常的高风险，高召回率是可以接受的，而不是损失。在图 8 中，我们展示了用细胞核“中级”数据集的两级微调的 ROC 性能。同样，我们注意到不管“高级”数据集大小，我们提出的两级逐步微调超过了最先进的单级微调方法。通过两级微调方案，当前最流行的深度神经网络的性能都能

达到相当于通过单级微调一个更大的好的标注的数据集的水平。

此外，如果我们将第一级微调中使用的两种类型的“中级”数据集进行比较，实际上几乎没有明显的差异。当我们有一个小的良好标注的数据集时，组织数据似乎更适合 VGG-16 和 Inception V3 模型，而 AlexNet 似乎更适合细胞数据。相应地，当我们有一个大的良好标注的数据集时，AlexNet 和 Inception V3 都得到了更有效的增强。因此，综上所述，所有这些结果都证明了基于组织或细胞数据集的逐步微调方法，成功地提高了用于各种情况下胃癌图像分类的预训练神经网络的性能。

TABLE 3: Performances of the proposed two-stage fine-tuning using cell-wise (nuclei) data.

Data size	Scheme	CNN architecture											
		VGG-16				AlexNet				GoogLeNet (Inception V3)			
		AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall	AUC	ACC	Precision	Recall
Small	One stage	0.879	0.793	0.863	0.695	0.828	0.723	0.74	0.72	0.838	0.753	0.75	0.75
	Two stage (nuclei)	0.902	0.815	0.873	0.730	0.852	0.777	0.78	0.78	0.872	0.784	0.78	0.78
Large	One stage	0.936	0.836	0.957	0.703	0.867	0.794	0.80	0.79	0.881	0.779	0.79	0.78
	Two stage (nuclei)	0.965	0.89	0.881	0.901	0.915	0.834	0.84	0.83	0.933	0.862	0.86	0.86

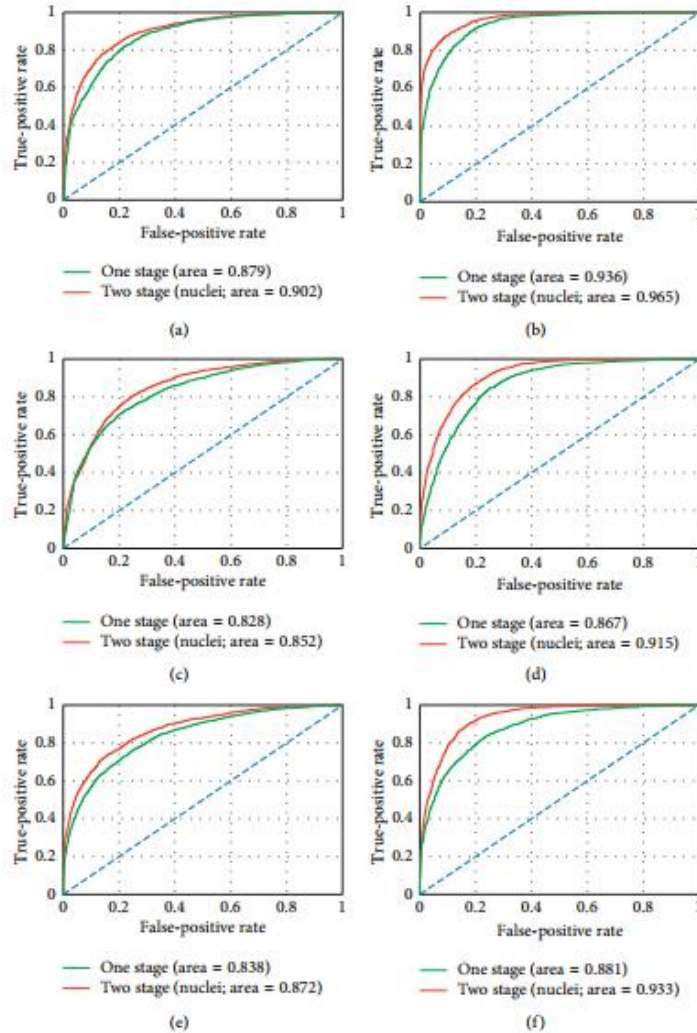


FIGURE 8: Performances of the proposed two-stage fine-tuning using cell-wise data presented by ROC. (a) VGG-16 + small data. (b) VGG-16 + big data. (c) AlexNet + small data. (d) AlexNet + big data. (e) Inception V3 + small data. (f) Inception V3 + big data.

5. 结论

本文针对病理图像分类训练数据标注不足的问题,提出了一种新的逐步微调方案和一种低成本的目标相关的中级数据,利用这些中间数据,深度神经网络可以根据病理医生的感知方式获取基本的病理知识。针对不同数据生产能力的需求,提出了两种实用的中间数据构建方法。这两种方法都不需要病理医生额外的工作量。在实验中,我们提出的方案针对不同的标注良好的病理图像数据,提高分类性能都发挥了足够的效能,并且对于不同的先进的 CNN 架构具有较高的应用能力。

最后,本文把最终的分类性能作为唯一的评价标准。客观地说,目标相关的中间数据应用于深度神经网络的有效性在为了的工作中有望得到具体的评价。此外,仍然有必要讨论使用微调的训练框架之间的差异一级其他有前途的技术,如多任务学习。我们希望这种使用低成本目标相关数据的训练方案能够为更多的医学图像分类提供解决方案。