基于深层卷积激活特征实现对组织病理学图像分类分割和可视化

Large scale tissue histopathology image classification, segmentation, and visualization via deep convolutional activation features

Yan Xu, ZhipengJia, Liang-Bo Wang, Yuqing Ai, Fang Zhang, Maode Lai and Eric-Chao Chang

目录

摘要

组织病理学图像的分析是癌症识别和诊断的金标准。病理图像的自动分析可以帮助病理学家诊断肿瘤和癌症的类型、并且可以缓解病理学家的工作量。数字组织病理图像分析有两个基本的任务：图像分类和图像分割。组织病理学图像阻碍自动分析的典型问题包括：复杂的临床表现、数据集中有限的训练图像的数量和大量的单数图像（通常上升到了亿像素）。即使数据集中的图像数量有限，但单个图像超大尺寸的特性也使得组织病理学图像数据集被认为是大规模的。

结果

本文利用深层卷积神经网络(CNN)激活特征来进行大规模组织病理学图像的分类、分割和可视化。我们的框架流程是从一个已训练的 CNNs 网络中提取的组织病理学图像特征，此外这个网络是由一个大型的自然图像数据库 ImageNet 训练的。我们还通过观察最后一个隐藏层中单个神经元组件的响应来探索 CNN 特征的特性。其中一些特征揭示了病理学家证实的生物学深刻见解。根据我们的实验，在 2014 年 MICCAI 脑瘤数字病理学挑战上大脑肿瘤数据集和结肠癌组织病理学图像数据集，该框架已经显示了最先进的性能。

结论

该框架是一种简单、高效、有效的组织病理学图像自动分析系统。我们成功地将 ImageNet 知识迁移作为深层卷积激活特性，用于训练数据少的组织病理学图像的分类和分割。显然 CNN 提取出的特征比专家设计的强大。

关键词

深层卷积激活特征，深度学习，特征学习，分割，分类。

1 背景

组织病理学图像分析是肿瘤识别和诊断的金标准[1,2]。数字组织病理学图像分析可以帮助病理学家诊断肿瘤和癌症亚型，减轻病理学家的工作量。在数字组织病理学图像分析中，有两种基本类型的任务:图像分类和图像分割。在分类任务中，算法以组织病理学全切片图像作为输入，并输出输入图像的标签。可能的标签是预先定义的，它们可以是特定类型的癌症或正常。在分割过程中，该算法将部分组织病理学图像作为输入，并对具有一定特征的输入图像进行区域分割。在这两个任务中，都给出了一组带有真实标签和注释的训练数据。在本文中，我们为所有这些相关的组织病理学问题(如分类和分割)建立了一个共同的框架，并通过可视化的方法来探索揭示关键生物学洞见的深层卷积激活特征的特性。

数字组织病理学图像的自动分析有三个主要的挑战:临床特征表现的复杂性，训练图像的数量不足，以及单个组织病理学图像的超大规模。

第一个挑战反映了复杂临床特征的困难。特征表示在医学图像分析中起着重要的作用[3,4]。不同癌症类型的组织病理学可以表现出不同的形态、规模、纹理和颜色分布，这使得我们很难找到一种适用于脑癌和结肠癌的肿瘤检测的一般模式。因此，特征表示[5]在分类和分割等高级医学图像任务中非常重要。很多以前的文献都集中在特征设计上，比如像类似目标（object-like）[6,7]和纹理特征[8,9]。然而，他们的设计的特殊性限制了一个固定的图像来源的应用。

另一个主要问题是医学图像领域的训练数据量不足。医学图像数据集通常比自然场景图像数据集小得多，这使得许多以前的机器学习算法不能直接应用于医学图像数据集。有两个因素使得收集医学图像的成本高昂。一是研究的疾病发病率低。由于图像的数量取决于疾病的发生率，研究的疾病的低频率使得采集过程更加困难。另一种是大量手动数据注释需要的劳动力，因为详细的医学图像手动标注通常需要大量的努力。此外，由于许多临床线索很难

量化，人工标记也具有内在的模糊性，即使被临床专家标记。

最后一个问题，单个组织病理学图像的巨大，使得组织病理学图像数据集被认为是大规模的;增加了计算复杂度,使图像分析更具挑战性。扫描一个典型的完整的组织病理学切片，可以产生一个超过 100000×100000 像素的图像，并包含 100 多万个描述性对象。通常情况下，在病理切片扫描过程中，每个病人可以获得 12 到 20 张图像。由于组织病理学图像数据集具有大规模属性，特征提取模型需要同时具有速度快和内存大的特性，而学习算法应该被设计为能够从这些大图像中提取尽可能多的置信。

在自动组织病理学图像分析的所有任务中都存在上述问题。除此之外，分类和分割任务也面临着一些具体的挑战。在分类中，不同癌症子类型之间的细微差别要求特征具有高度的表达性。不同子类型的不平衡实例也妨碍了分类器的使用。在分割任务中，区域的定义不同，这使得由多个病理学家标注的真实值略有不同。这种模糊属性在分割框架的设计中成为一种挑战。

随着深度卷积神经网络(CNN)的出现，CNN 的激活特征最近在计算机视觉上取得了巨大的成功[10-16]。像 ImageNet 这样的大型可视化数据库的出现，其包括超过 1000 万张图片和超过 20,000 个类[17]，使得 CNNs 能够从一般图像中提供丰富多样的特征描述。CNN 隐藏层的响应提供了不同层次的图像抽象，可以用来提取复杂的特征，如人脸和自然场景。它使得从医学图像中提取足够的置信成为可能。因此，本文利用深度卷积激活的方法研究了 ImageNet 知识的潜力，提取了组织病理学图像分类和分割的特征。

虽然 CNN 本身可以进行图像分类[14]和分割[18]，但是单个组织病理学图像的超大规模使其直接对 CNN 进行分类或分割是不现实的。一方面，构建一个非常大的输入尺寸的 CNN 是不实际的。另一方面，将整个组织病理学图像缩小到 CNN 可以接受的大小，将会丢失太多的细节置信，这使得它不可能被识别，即使是病理学家。基于这一事实，我们的分类和分割框架利用 CNN 的激活特性，采用了一种小块采样技术，这些局部小块的局部细节将被保留下来。最后的结果采用了不同的策略。在分类框架中，使用特征池来构造所有全切片图像的特征。在分割框架中，在小块层次上进行分类，并通过结果来构造图像范围的分割。更小的小块和平滑可以使边界更精确。

为了使 CNN 的激活特性更适合于组织病理学图像，我们还对 ImageNet 模型进行了微调，以了解更微妙和更有深度的特征，以捕获复杂的临床表征特征。在我们的实验中，微调的 CNN 模型可以在分类和分割任务上达到更好的精度。

此外，我们通过对组织病理学图像分类中 4096 维特征向量的可视化，探索了 CNN 激活特征的特性。计算了每个图像的小块中心的热图和 CNN 激活特征的单个神经元的识别块。热图解释了哪些小块或区域产生了强烈的反应，使它们的图像落在相应的类别中，而代表单个神经元反应的小块帮助我们理解这些反应从每个分类器的角度所具有的特征。通过这个可视化分析，我们发现了临床知识和我们的方法的反应之间的一些关系。

本文提出了一种简单、高效、有效的方法，利用 CNN 的激活特征对组织病理学图像进行分类和分割。通过实验，我们的框架在两个数据集中取得了良好的性能。我们的框架的优点包括:

1. 将 ImageNet 训练的 CNN 获得的强大的特征，转移到组织病理学图像中，解决了组织病理学图像数据集中训练数据有限的问题;

2. 采用小块采样和池化技术，利用本地描述性的 CNN 特性，使整个框架具有可扩展性和效率，在非常大的全切片组织病理学图像中;

3. 对于两种不同癌症类型，使用了统一框架，表明了我们方法的简单性和有效性。

我们对组织病理学图像的自动分析领域做出了两项贡献:

1. 一种多用途的病理组织学问题的解决方法，在两种不同类型的癌症上证明有效;

2. 一个可视化的策略揭示了我们的框架所学习到的特征具有生物学的洞察力，并证明了 CNN 的激活特征在表征复杂的临床特征方面的能力。

徐等人提出了与我们的方法类似的早期会议版本[19]。在本文中，我们进一步说明:(1)框架方法可用于分析脑肿瘤以外的组织类型，如结肠癌；(2)添加基于 ImageNet 模型的微调功能；(3)引入热图，探究在分类任务中，哪些小块或区域在一个图像中提供了强烈的反应，同时伴随了个体神经反应的可视化。

2 相关工作

近年来，数字组织病理学的应用呈现出巨大的发展趋势。研究人员试图用数字组织病理学取代光学显微镜作为病理学家使用的主要工具。在[20-23]中研究了各种替代方法。在研究数字组织病理学的趋势下,多次举行提高肿瘤组织病理学研究的比赛，包括 2012 年 ICPR 有丝分裂检测竞争[24]，2013 年 MICCAI 挑战赛在有丝分裂检测[25]，2014 年 MICCAI 脑瘤数字病理挑战[26]，2015 年 MICCAI 腺分割挑战竞赛[27]。我们提出的框架在 2014 年 MICCAI 脑瘤数字病理学挑战中，首先取得了分类和分割的结果[28]。

特征表示设计是与组织病理学图像相关的一个重要方向。手动设计的特征包括分形特征[29]、形态测量特征[30]、纹理特征[31]和目标特征[32]。Kalkan[33,34]利用小块层级的图像纹理特征和结构特征，提出了一种区分结肠癌和非癌症的两级分类方案。Chang[35]在不同的位置和尺度上提出了稀疏组织形态学特征，以区分 GBM 数据集下的肿瘤、坏死和转化为肿瘤的细胞，KRIC 数据集下的正常和基质细胞。由于数据量大，Chang 也使用了空间金字塔匹配来表示多尺度特征。Rashid[36]设计了两种特殊的腺体特征，用于描述前列腺腺癌的良性和恶性腺体。这两个特征是核层数和上皮层面积与腔面积的比值。Song[37]用基于学习的过滤器转换图像，以获得更具有代表性的特征描述符。Sparks[38]提出了一组新的明确的形状特征，以区分前列腺腺体与前列腺癌之间细微的形状差异。Sos Agaian[39]介绍了组织描述的新特征，如超复小波分析、四元数色比和局部改变模式。

然而，这些方法的主要问题是选择鉴别特征来表现临床特征的困难。研究[40]还表明，由双层网络学习的特征比手动设计的组织病理学图像更强大。Nayak[41]利用受限玻尔兹曼机(RBM)对稀疏特征学习进行了探索，以描述清晰细胞肾癌(KIRC)和 GBM 的组织病理学特征。这些研究表明，特征学习优于特殊的特征设计。但是，在特征学习中存在一个普遍的挑战，即在许多情况下，训练数据的数量是有限的。在我们的案例中，只有一些训练图像可以用于分类和分割。

在许多医学图像任务中，使用深层 CNN 特性作为通用表示是一种日益增长的趋势。一些公共可用的深层 CNN 模型被用来提取特征：Caffe[42]在许多作品中被开发[10,11,42]和 OverFeat[43]被[16]使用。这些特征通常用于分类和对象检测任务[10,11,16,42]。然而，这些研究只关注自然图像。

强大的 CNN 不仅能够进行分类，而且能够学习特征，并且有几项研究直接利用 CNN 的这一属性进行组织病理学图像分析。Ciresan[24]修改了传统的 CNN，使之成为一个深度融合的 CNN，以检测乳房组织学图像中的有丝分裂。将检测问题作为像素分类。以像素为中心的小块中的置信作为上下文使用。他们的方法在 2012 年 ICPR 有丝分裂检测比赛中获得了第一名。训练集只包括 5 个不同的活检 H&S 染色全切片，其中包含大约 300 个全丝分裂事件。Cruz-Roa[44]为自动基底细胞癌检测提供了一种新的深度学习架构。训练集包含了来自 308 个皮肤组织病理切片的 1417 张图片。相比之下，ImageNet[17]由大约 1400 万张图片组成，比组织病理学图像的数据集要大得多。基于我们对特征设计和特征学习的调查，我们决定采用 ImageNet 训练的 CNN 特征来描述脑肿瘤和结肠癌组织病理学图像的鉴别纹理。

微调是 CNN 学习的一个重要步骤。它维护原始的网络架构，并将训练过的 CNN 作为初始化。经过微调训练后，新模型可以学习更精细的表示来描述新的目标任务。Ross[45]在

2007 年的 VOC 测试中，提出了利用微调来提高 10%的目标，从 44.7%(R-CNN fc7)到 54.2%(R-CNN 微调 fc7)。张[46]给出了一个细粒度的分类。通过使用微调特性，使用预先训练的 CNN 特性，准确度从 68.07%提高到 76.34%。这些研究表明，微调是有效的。在我们的例子中，在预先训练的 CNN 特性的基础上，我们实现了微调步骤，以学习更多关于组织病理学图像的精细表示。

除了特征表征之外，组织病理学图像分析还涉及到分类方案。Xu[47,48]引入了一种新的模型，称为多聚类实例学习，其表现为拓扑癌症图像分类、分割和聚类。此外，Xu[49]提出了上下文约束的多实例学习来采用分割。Gorelick[50]提出了一个基于两个阶段的 AdaBoost 分类。第一阶段识别组织成分，第二阶段使用公认的组织成分来分类癌和非癌性，高等级和低级别癌症。Kandemir[51]引入了一个概率分类器，它结合了多个实例学习和关系学习来区分癌和非癌。分类器利用图像级置信和不同癌症状态下的细胞形态变化。Kalkan[33]提出了两阶段分类。第一阶段将小块分类为可能的类别(腺瘤性、炎症性、癌性和正常性)。第二阶段使用第一阶段的结果作为特征。最后，一个逻辑的线性分类器识别癌症和非癌症。在我们的例子中，一个线性的 SVM 分类器被用来考虑它的简单和快速。

在分类中，所使用的输入通常是调整后的原始图像[14]。所提取的 CNN 特征被直接用作分类的最后特征。[14]有几种不同的方法。Sharif Razavian 等[16]提取了 16 个小块，其中包括原始图像、5 种作物（crops）(原图像区域 4/9 的四个角和一个中心)，以及两个旋转和它们的镜像。当使用 16 个小块作为输入时，将提取 CNN 特性。在此之后，作者[16]将最后一层的所有响应的总和作为最终的特征。Gong 等[11]以 32 个像素的跨度，在多尺度水平上对小块进行采样。提取了深层卷积激活特征的多尺度有序集。然后，作者[11]通过本地聚合描述(VLAD)编码的矢量来聚合本地小块响应。在我们的方法中，受到[52]的启发，并且观察到组织的图像与图像的十亿像素大小非常接近，我们使用小块采样来生成许多小块来保护详细的本地置信，并使用特征池来将小块级别的 CNN 特性聚合到最后的特征中。

组织病理学图像分析应用于广泛的研究领域。Khan[53]提出了一种非线性映射方法，使染色法标准化。当不同的组织制备、染色反应、用户或协议以及不同厂家的扫描仪被使用时，应用特定于图像的颜色反褶积处理颜色变化。Zhu[54]提出了一种新颖的批模式主动学习方法，以解决可扩展组织病理学图像分析中标注的挑战。特征选择和特征减少方案[38,55]也是组织病理学图像分析的重要步骤。

3 方法

3.1 CNN 结构

AlexNet[14]是一种简单而常见的深层卷积神经网络，与其他类型的网络相比，它在分类上仍能取得较好的竞争性能。因此，在我们的案例中使用了 AlexNet 体系结构。我们在本文中使用的 CNN 模型由 ImageNet LSVRC 2013[13]的 CognitiveVision 团队共享[0]，其架构如表 1 所示。

Table 1 The CNN architecture

Layer	Dimension	Kernel size	Stride	Details
input	224 × 224 × 3	-	-	RGB channels
conv1	55 × 55 × 96	11	4	-
pool1	27 × 27 × 96	3	2	Max pooling
conv2	27 × 27 × 256	5	1	-
pool2	13 × 13 × 256	3	2	Max pooling
conv3	13 × 13 × 384	3	1	-
conv4	13 × 13 × 384	3	1	-
conv5	13 × 13 × 256	3	1	-
pool3	6 × 6 × 256	3	2	Max pooling
fc1	4096	-	-	-
fc2	4096	-	-	-

它类似于[14]中使用的，但是没有 GPU 拆分，因为一个现代 GPU 有足够的内存用于整个模型。这个模型在整个 ImageNet 数据集上进行了训练。因此，它与在 2013 年 ILSVRC 年使用的认知视觉团队没什么不同。用于训练和提取特性的代码基于[14]。在训练步骤中，我们使用[14]中引入的数据预处理和数据增强方法，将各种分辨率的输入图像转换为 224×224。在特征提取过程中,输入图像的大小为 224×224 像素和提供到网络中。将 fc2 层的输出作为提取的特征向量。

3.2 分类框架

组织病理学图像的巨大规模使其在局部提取特征变得单一。因此,每个组织病理学图像分为一组重叠的正方形小块大小为 336×336 像素放大 20×和 672×672 像素放大 40×(他们都是 151872×151872 平方厘米)。这些小块组成一个矩形网格，有 64 像素的步幅，即，相邻块之间的距离。为了进一步减少小块的数量，我们丢弃了只有白色背景的小块，所有像素的 RGB 值都大于 200。所有选定的小块都被调整到 224×224 像素，并输入网络以获得 4096 维的 CNN 特征向量。图像的最终特征向量是通过 P-范数池计算的。P-范数池，也称为 softmax 汇聚，从几个小块中放大信号，这是经过计算的。

$$f_P(\mathbf{v}) = \left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{v}_i^P\right)\frac{1}{P}, \qquad (1)$$

其中 N 为图像的小块数，$\mathbf{v_i}$ 是第 i 个小块特征向量。在我们的框架中，使用 P=3(3-范树池)。

此外，为了形成更具有鉴别特征的子集，并排除冗余或不相关的特征，使用二分类分类进行特征选择。特征是根据阳性和阴性标签的区别来选择的。第 k 个特征差异 $diff_k$ 计算如下：

$$diff_k = \left|\frac{1}{N_{\text{pos}}}\sum_{i\in\text{pos}} v_{i,k} - \frac{1}{N_{\text{neg}}}\sum_{i\in\text{neg}} v_{i,k}\right|, \qquad (2)$$

其中 k=1，…，4096，N_{pos} 和 N_{neg} 是训练集中正负图像的个数，$v_{i,k}$ 是第 i 个图像的 k 维特征。然后将特征组件从最大的 $diff_k$ 排序到最小的，然后选择前 100 个特征组件。对于多类分类，没有使用特征选择。

最后，采用线性支持向量机(SVM)。在多类分类中，使用了 one-vs-rest 分类。图 1 显示了我们的分类框架的流程。

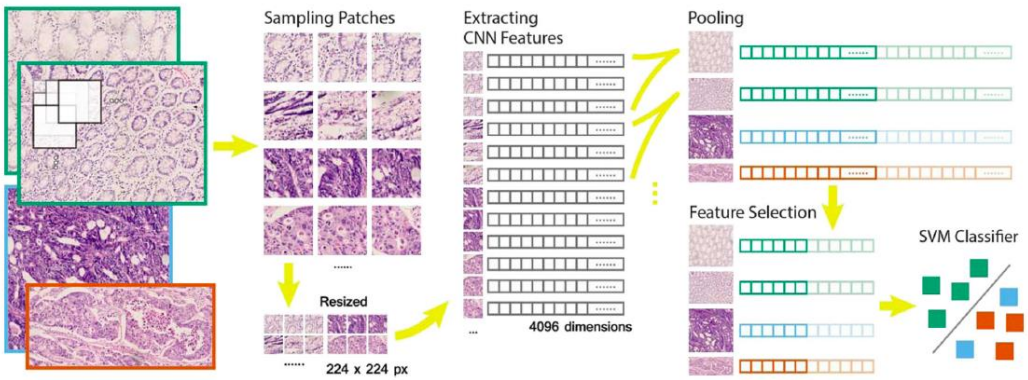

图 1. 分类工作流程。首先，根据图像的放大倍数，在矩形网格上采样大小为 336 或 672 像素的正方形块。然后将小块大小调整为 224 像素，作为 CNN 模型的输入。每个小块从 CNN 模型中提取一个 4096 维的特征向量。每个图像的特征池和特征选择获得了一个 100 维特征。最后，线性支持向量机对所选特征进行分类。图中显示的是一个二元分类，其中阳

性(蓝色和橙色)和阴性(绿色)分别是脑瘤的 GBM 和 LGG。在多类分类中，使用了 4096 维的全特征向量。

3.3 分割框架

医学图像分割方法一般可分为三类:监督学习[29]、弱监督[48]、和无监督[32]。有监督的学习方法只能在有标签的数据可用的情况下使用。否则，需要其他方法(即非监督方法)。由于我们已经标记了训练数据，所以我们提出了一个监督学习框架来进行分割。在我们的框架中，我们通过对一个小块集合进行分类，将分割问题重新定义为一个分类。图 2 说明了分割框架的流程。

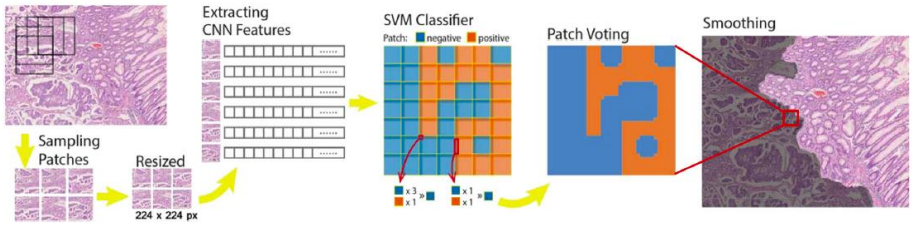

图 2. 分割工作流程。类似于分类工作流程，在矩形网格上以 8 像素的步幅对 112 像素大小的正方形块进行采样。每个小块都有一个阳性(橙色)或阴性(蓝色)标签，分别是脑瘤的坏死和非坏死，以及非正常结肠癌和正常结肠癌。在训练阶段，如果小块与带注释的分段区域重叠比大于 0.6，则小块为阳性。然后对小块进行调整，从我们的 CNN 模型中提取一个 4096 维的特征向量。用一个线性 SVM 分类器来区分阳性和阴性。利用所有预测的置信值，得到概率映射图像。经过平滑处理，得到了正分段。

类似于前面提到的分类框架，小块以 8 个像素为步长，在一个 112×112 像素小块的矩形网格中采样。调整 112×112 像素块大小为 224×224 像素来获取他们的 CNN 特征向量。使用所有的小块对一个线性支持向量机训练。由于像素可以被多个不同标签的重叠小块覆盖，因此每个像素的最终标签由覆盖该像素的小块的多数投票决定。由于基于像素的投票提供了许多缺乏生物学意义的微小的阳性或阴性的区域，我们利用几种平滑技术来减少区域分数。面积小于全部图像大小的 5%的阳性和阴性区域移除。

在 MICCAI 挑战中，我们进一步对最终提交模型的训练数据进行了两次修改。

1. 我们观察到出血组织出现在非坏死和坏死区域。因此，我们在坏死区域中手动地将出血小块作为非坏死小块。这导致在预测阶段的测试时间出现了失血小块的分类错误，但由于这些小块通常是在坏死区域的内部，这样的错误可以通过后处理来纠正。

2. 我们观察到训练图像是不均匀的，并且有不同的尺寸。另外，训练数据分布不均匀。在最后的提交模型中，我们通过对训练数据进行交叉验证，增加了遗漏的区域和错误区域的实例。

3.4 数据集

我们将我们的分类框架和分割框架用于两个组织病理学图像数据集: 2014 年 MICCAI 脑瘤数字病理学挑战和结肠癌数据集。为了说明我们的框架的优点，我们还在相同的数据集上测试了其他方法和其他类型的特性。

对于 MICCAI 的挑战[26]，组织方提供了脑肿瘤的数字组织病理学图像数据。在分类(次挑战 I)中，目标是区分胶质母细胞瘤的图像(GBM)和低级别胶质瘤(LGG)。训练集有 22 个 LGG 图像和 23 个 GBM 图像，测试集有 40 个图像。在细分(次挑战 II)中，目标是将坏死和非坏死区域从 GBM 组织病理学图像中分离出来，因为坏死是区分 LGG 和 GBM 的重要线索。训练集包括 35 个图像，测试集包含 21 个图像。图像分辨率分别为 502nm/pixel 或 226nm/pixel，分别对应 20×和 40×的镜头放大。

对于结肠癌，H&E 染色组织病理学图像由中国浙江大学病理学系提供，并由滨松的

NanoZoomer 幻灯片扫描仪扫描。在三个组织病理学家的审查过程中，对包含典型癌症亚型特征的区域进行了剪切和选择，其中两个病理学家独立地提供了他们的结果，而第三个病理学家在他们的注释中合并和解决冲突。共有 717 个裁剪区域被用作我们的数据集,最大的规模 8.51×5.66 毫米,平均面积 5.10 平方毫米。所有图像均为 40×倍放大倍数，即 226 nm/pixel。355 个癌症和 362 个正常图像用于二分类任务。对于多类分类，有 362 例正常(N)，154 例腺癌(AC)，44 例黏液癌(MC)，50 例锯齿状癌(SC)，38 例乳头状癌(PC)，45 例筛管型腺癌(CCTA)图像(共使用 693 张图像)。癌症的图像被忽视，因为在他们的癌症类别中有太少的实例。一半的图像被选择作为训练数据和其他图像作为测试数据。测试数据中每种癌症子类型的比例与完整数据集相同。在分割任务中，从数据集中选择 150 个训练和 150 个测试图像。他们的大小 10×放大规模(904nm/pixel),然后裁剪 1280×800 像素。这是在[32]中为他们的算法 GraphRLM 所使用的相同设置。病理学家对结肠癌图像分割的基本事实进行了注释，遵循了之前提到的相同的审查过程。

4 实验设置

4.1 分类

为了说明 CNN 特征的优点，我们在我们的框架中比较了 CNN 提取的特征和手动提取的特征(有固定的提取算法)。仅对框架中的特征提取步骤进行了修改。在我们的实验中，采用了包括 SIFT、LBP 和 L*a*b 颜色直方图的通用对象识别功能(在[48]中进行了如下设置)，并将其连接到 186 个特征维度中。该方法由 SVM-MF 表示，我们所建议的使用 CNN 特征的框架由 SVM-CNN 表示。

为了显示小块采样的有效性，我们将我们的框架与直接使用 CNN 特征的方法进行比较，而不使用小块采样。在此方法中，完整的组织病理学图像被调整为 224×224 像素，并由 CNN 提供给 CNN 以提取图像级特征。然后用线性支持向量机进行分类。该方法由 SVM-IMG 表示。

此外，我们还比较了我们的分类框架和前面的方法[48]和判别数据变换[37]。它们分别用 MCIL 和 TRANS 表示。在 MCIL 中，小块提取设置与我们的方法相同。这里的 softmax 函数是 GM 模型，弱分类器是高斯函数。该算法的参数与原始研究中描述的相同。在反式中，基于学习的过滤器应用于原始图像和特征描述符[37]。我们在他们的原始工作中遵循设置(X=3、5、7 和 Y=5 的特征过滤器的图像过滤器)，并使用一个线性 SVM 作为分类器。

在所有的方法中都采用了线性 SVM(SVM-IMG、SVM-MF、SVM-CNN 和 TRANS)，在实验中采用了 L2 正则化的 SVM，其成本函数为 $\frac{1}{2}w^Tw + \dot{C}\sum_{i=1}^{l}$。使用开源工具箱 LIBLINEAR[56]来优化 SVM。参数 C 的取值从{0.01,0.1,1,10,100}中选出，最优值是通过对训练数据的交叉验证来确定的。

4.2 分割

与分类类似，我们将 CNN 获得的特征与手动获得的特征进行比较。手动提取特征设置与分类实验相同。该方法由 SVM-MF 表示，我们所建议的使用 CNN 特征的框架由 SVM-CNN 表示。

为了进一步提高分割结果，由 ImageNet 训练的 CNN 模型对组织病理学图像进行了微调，以探索更适合这项任务的特征。在我们的实验中，我们使用随机的初始化 2 类的分类层代替原来的 1000 类的分类层。CNN 的架构保持不变。我们以 0.0001 的学习速率开始一个随机梯度下降(SGD)。在未修改的图层中使用学习速率，这是 ImageNet 初始训练率的十分之一。我们训练了 20 次迭代的 CNN 模型，在训练过程中学习率没有下降。除了从微调的 CNN 模型中提取的特征之外，分割框架的其他步骤也不会改变。该方法由 SVM-FT 表示。

此外，我们将我们的分割框架与前面的方法 GraphRLM[32]进行了比较。由于我们的数据集和原始数据集都是相同放大倍数的结肠癌数据集，所以我们的实验中的参数设置与发布时相同 $r_{\min}=8, r_{\text{strel}}=2, win_{\text{size}}=96, dist_{\text{thr}}=1.25$和 $comp_{\text{thr}}=100$。这个方法用 GraphRLM 表示。线性支持向量机的设置与分类实验相同。

4.3 评估

对于分类任务，准确性被用作评价得分。对于分割任务，评估遵循 MICCAI 挑战的组织者提供的规则，该规则计算每个图像的重叠区域大小的平均比，与算法所预测的真实值和结果的总面积大小有关。分数的计算如下。一个映射定义了一组被分配给一个阳性标签的图像的像素。让第 i 个图像分割的真实值映射为 G_i，算法生成的映射为 P_i。图像 i 的分数为 S_i，计算如下：

$$S_i = \frac{2\,|P_i \cap G_i|}{|P_i \cup G_i|},\ i = 1, \ldots, K, \qquad (3)$$

K 是图像总数。 S_i是评估分数（准确率）。

对于脑瘤任务，由于 MICCAI 挑战的组织方没有提供真实标签和测试数据的注释，我们在实验中使用 5 折交叉验证进行分类和交叉验证。另外，第 2.3 部分中提到的修改并不适用于我们自己的交叉验证实验。

5 结果和讨论

5.1 分类结果

在 MICCAI 挑战赛中，我们最终提交的分类任务对测试数据的准确性达到了 97.5%，在其他参与者中排名第一。表 2 显示了提交网站上提供的一些最优秀的方法的结果[28]。

Table 2 Classification performance in the MICCAI challenge

	Accuracy	Place
Anne Martel	75.0%	4th
Hang Chang [30]	85.0%	3rd
Jocelyn Barker	90.0%	2nd
Our method [19]	97.5%	1st

我们的结果是令人满意的，我们的性能和排名第二的团队之间的差距达到了 7.5%，这证明了我们的方法在 ImageNet 的帮助下可以达到最先进的水平，即使是相对较小的数据大小。

从 MICCAI 的挑战中获得数据，我们将我们的方法与先进的技术方法相比较。表 3 总结了一些先进技术的表现。

Table 3 Classification performance using cross-validation in training data from the MICCAI challenge

	Accuracy
Hang Chang [30]	85.83%
Our method [19]	97.8%
Jocelyn Barker [57]	100.0%

我们的结果优于其他方法。方法[57]采用两阶段、粗到细的分析，大大减少了计算时间，比任何实时应用程序都要慢。我们使用 NVIDIA K20 GPU 来训练我们的模型。该挑战图像

的坏死和非坏死的像素分别为 1330,000 和 2,900,000。在测试时，使用我们的滑动窗口方法来预测整个图像分割的平均计算时间是 GPU 上的第二等级。

添加了我们的结肠数据集和多类分类场景，我们分别在脑瘤和结肠癌数据集上比较了几种方法。这些性能展示见表 4。

Table 4 Classification methods comparison

Dataset	MCIL	TRANS	SVM-IMG	SVM-MF	SVM-CNN
MICCAI brain	91.1%	86.7%	62.2%	77.8%	97.8%
CRC binary	95.5%	92.3%	94.3%	90.1%	98.0%
CRC multiclass	-	78.5%	79.0%	75.5%	87.2%

由于算法的局限性，MCIL 被排除在多类分类比较中。在所有情况下，我们的方法(SVM-CNN)的结果都有统计学意义。

对于 GBM 和 LGG 类型的脑瘤分类，CNN 提取的特征比手动特征强多了，性能提高了 20.0%。与 MCIL 和 TRANS 相比，我们提出的框架分别是 6.7%和 9.1%。

对于结肠癌的二分类，虽然我们的方法获得了与脑瘤相似的最高表现，但所有方法都至少达到了 90%的准确率。在多类场景中，只有我们的方法达到了 80%以上的精度。与其他方法相比，SVM-CNN 在使用完整图像时，直接使用了 8.2%，并打败了使用硬编码的手动特性的 SVM-MF。令人惊讶的是，在结肠癌中，SVM-IMG 比 SVM-MF 的表现要好约 4%。

在二分类分类中，MCIL 和 SVM-CNN 都比其他方法获得了更好的性能。由于 MCIL 是一种多实例学习算法，而我们的框架采用特征池技术，这与多实例学习相似，主要的性能差异是由强大的 CNN 特性贡献的。使用在一般图像数据库上训练的提取特征，即使训练图像的数量有限，我们也能捕获复杂和抽象的模式。

为了更好地捕捉在我们的组织病理学图像分析方法中被激活的特征，绘制了图像层级的热图(图 5 和图 6)和特征小块特征(图 7 和图 8)。

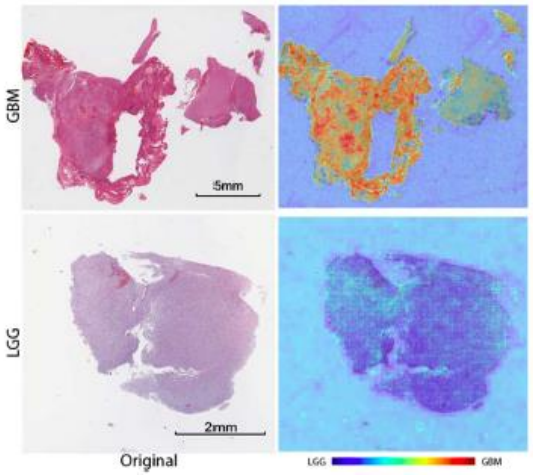

图 5.脑瘤 GBM 与 LGG 分类的热图。整个全切片图像的每个小块都使用分类器来分配一个信心，这个分类器形成了热图。红色的区域更可能是 GBM 区域。这些热图的目的是为了说明整个全切片图像的哪一部分对于分类器来说是重要的，并证明了 CNN 的表现力。在 GBM 的例子中，被认为是 GBM 诊断的重要形态学线索的内皮细胞增生区域表现出高度的积极自信。

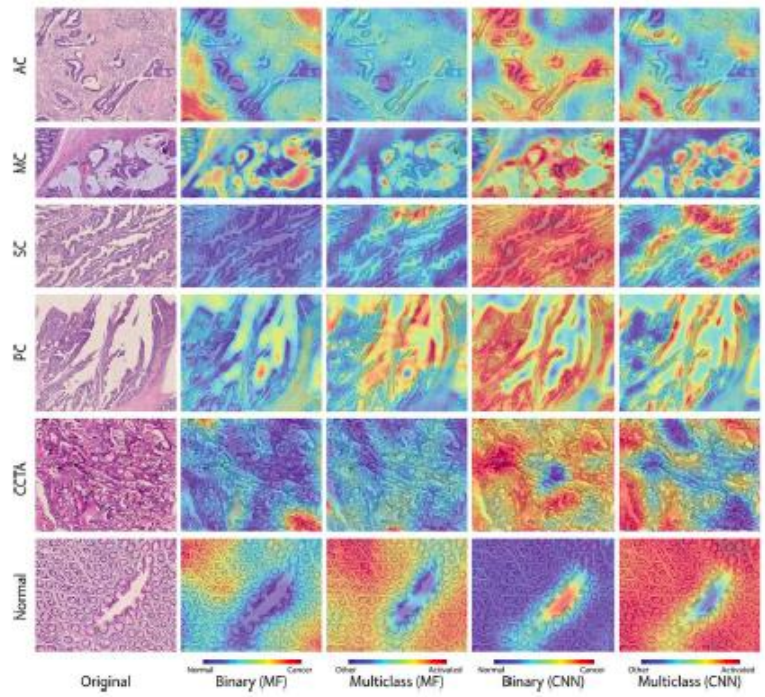

图 6. 利用手动提取的特征和 CNN 激活特征，对结肠癌的二分类和多类分类的热图。如图 5 所示，基于每个小块的置信得分，绘制了热图，目的也是为了探究 CNN 特色的表现力。在二分类(第 2 和第 4 列)中，红色区域更可能是癌症。在多类分类(第 3 和第 5 列)中，只显示预测图像标签的分类器，即对于 AC 图像，只显示 AC-vs-rest 分类器的预测。红色区域更可能是图像的标签。从二分类到多类分类的突出显示区域的转换表明，我们的多类分类器能够识别每种癌症子类型的特定特征。CNN 的特点和手动特征的比较显示了 CNN。特性比手动特性具有更大的表现力。

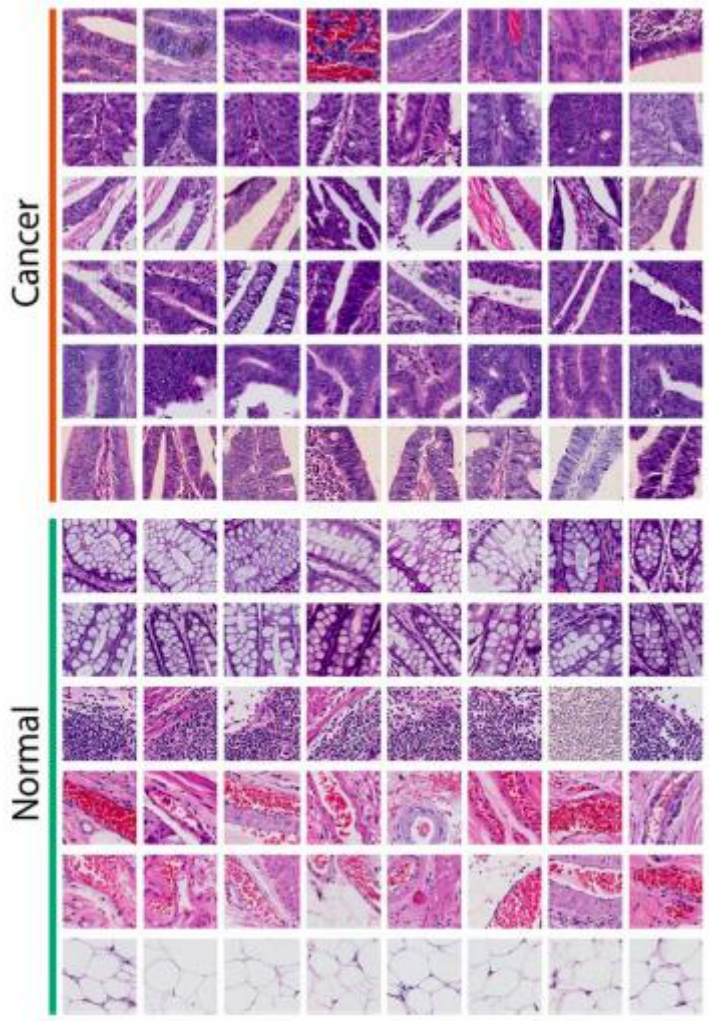

图 7. 在 CNN 激活特征的单个组件(神经元)中选择的样本鉴别小块。每一排的小块都在一个 4096 个神经元中引起一个高的响应，从所有的结肠训练图像的二进制分类任务。选择每个分类器的 6 个最重要的特征，并选择触发这 6 个神经元的顶部小块来表示相应特征。这个图的目的是展示 CNN 特征的各个组成部分的特征，这些特征被二进制分类器认为是重要的。这些可视化的特征传达了一些临床的见解。

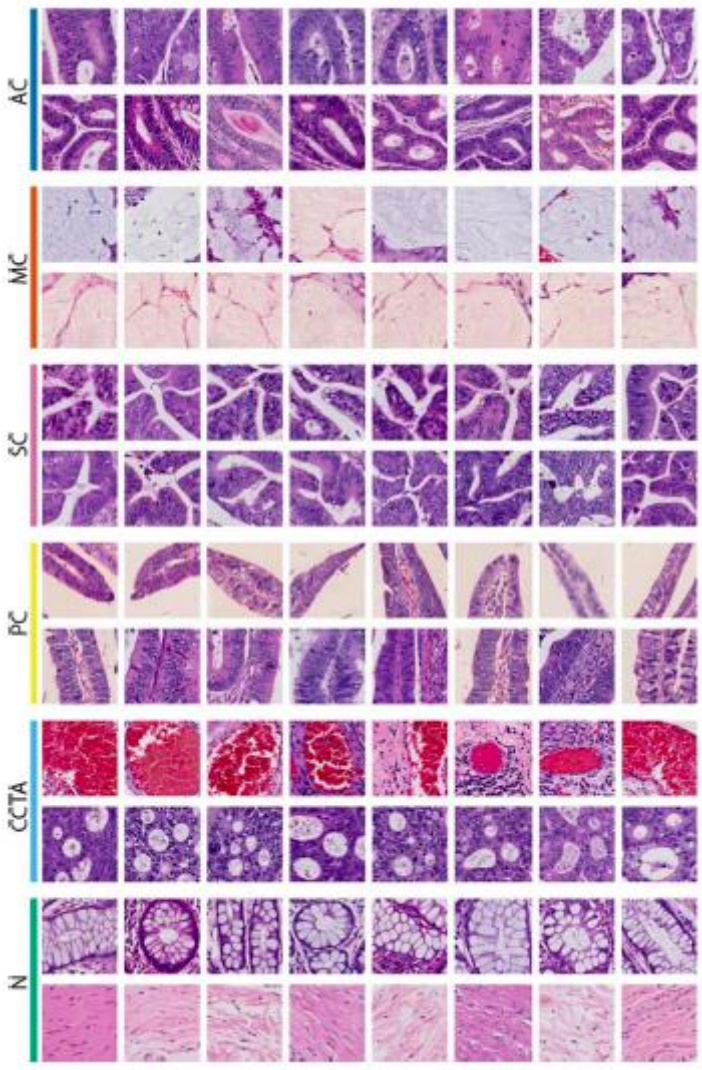

图 8. 在 CNN 激活特征的单个组件(神经元)中选择的样本鉴别小块。每一排的小块在一个 4096 个神经元中引起一个高的响应，从所有的结肠训练图像的多类分类任务。选择每个分类器的两个最重要的特征，并选择触发这两个神经元的顶部小块来表示相应特征。这个图的目的是为了显示 CNN 特征的各个组成部分的特征，这些特征被多类分类器认为是重要的。这些可视化的特征传达了一些临床的见解。

5.2 分割结果

在 MICCAI 的挑战中，我们的最终的分割提交也获得了第一个位置，在测试数据上的准确性高达 84%。表 5 显示了其他参赛队伍的表现[28]。我们的框架比第二名的团队要高出 11%。

Table 5 Segmentation performance in the MICCAI challenge

	Accuracy	Place
Anne Martel	63%	4th
Hang Chang	68%	3rd
Siyamalan Manivannan [58]	73%	2nd
Our method [19]	84%	1st

表 6 总结了不同方法对脑瘤和结肠癌数据集的分割效果。由于它是一种无监督的方法，所以 GraphRLM 不适合在这里进行比较。对于大脑肿瘤数据集，SVM-CNN 在 SVM-MF 上的性能提升了 21.0%。使用微调的 CNN 进一步提高了 SVM-CNN 的 0.4%。

Table 6 Segmentation methods comparison

Dataset	GraphRLM[1]	SVM-MF	SVM-CNN	SVM-FT
MICCAI brain	-	64.0%	84.0%	84.4%
CRC	-	77.0%	93.2%	94.8%

GraphRLM is an unsupervised method

对于结肠癌，基于 CNN 的方法在 SVM-MF 上至少显示了 16.2%的性能改善，所以结果表明与脑癌数据集有相似的趋势。经过微调，准确度进一步提高到 94.8%，显示出 1.6%的差异。此外，我们还提供了一些使用所有方法的分割结果的样本，分别如图 3 和 4 所示，分别用于脑瘤和结肠癌数据集。

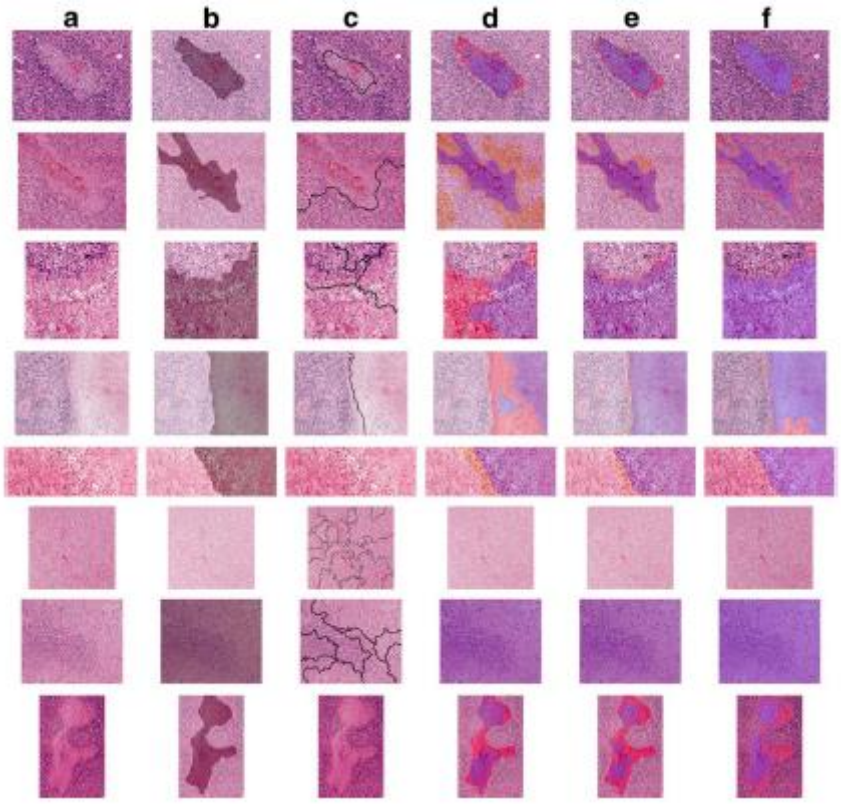

图 3. 脑肿瘤数据集的分割结果。a.一个原始的图像；b.坏死(阳性)区域的真实标记，灰色区域。其余的列显示了 c GraphRLM、d SVM-MF、e SVM-CNN 和 f SVM-FT 方法的预测结果，其中，true positive、false positive(missed)、false negative(错误预测)区域分别为紫色、浅红色和橙色。

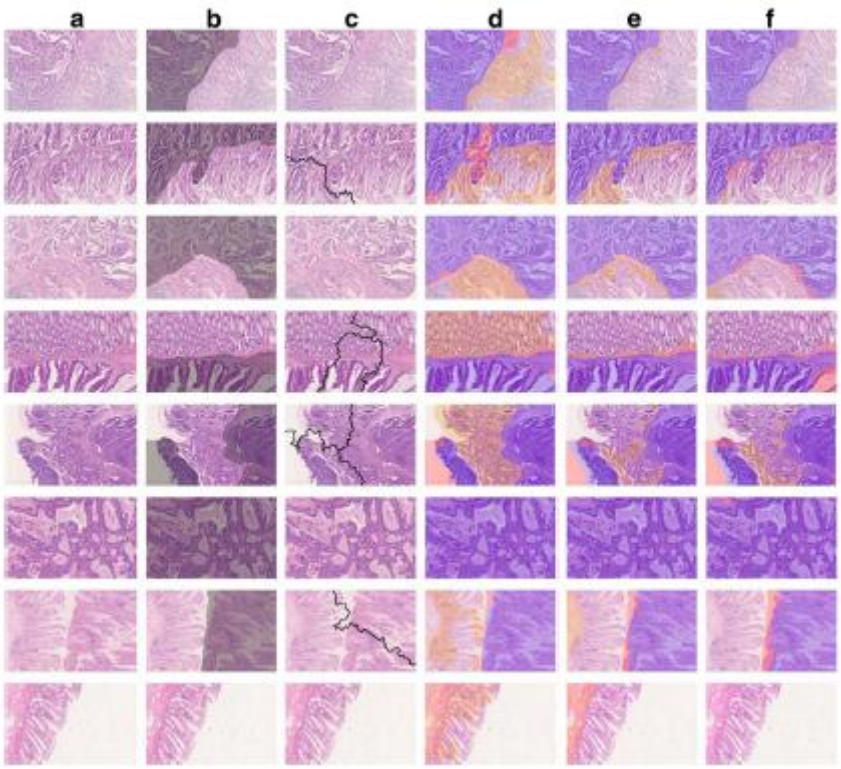

图 4. 结肠癌数据集的分割方法比较。a.一个原始的图像。b 坏死(阳性)区域的真实区域表示为灰色。其余的列显示了 c GraphRLM、d SVM-MF、e SVM-CNN 和 f SVM-FT 方法的预测结果。阳性(漏失)，假阴性(错误预测)区域分别为紫色、浅红色和橙色。

从表 6 中可以看出，使用基于 CNN 的特性而不是手动硬编码（hard-coded）的特性可以看到显著的性能差异。使用微调的 CNN，在结肠癌中提高了 1%的 CNN 特征的准确性。图 3 和图 4 的差异也可以得到验证。对于 GraphRLM，分割结果是无法求证的，或者没有提供分割结果。虽然 GraphRLM 的结果不能精确地量化，但它不能勾勒出有价值的边界，或者在大多数情况下不会产生边界。即使在结肠癌中，在他们的出版物中使用的同样的癌症类型，GraphRLM 也不能提供类似的形态模式的分割。另一方面，所有其他方法至少达到 64%的准确度。SVM-CNN 和 SVM-FT 在精度统计和可视化方面的性能都有明显的改善。

5.3 小块大小的选择

在我们的分类框架,采样小块的大小分别是 336×336 像素 20×放大和 672×672 像素 40×放大。我们也尝试其他的小块尺寸来探索不同小块尺寸的影响。结果如表 7 所示。从结果在表 7 中,我们发现一块大小为 672×672 收益率最高的准确性二分类和多级分类任务。

Table 7 Classification results of different patch sizes

Dataset	224×224	448×448	672×672
MICCAI brain	91.1%	93.3%	97.8%
CRC binary	97.5%	96.9%	98.0%
CRC multiclass	85.0%	85.3%	87.2%

Patch sizes in the table correspond to 40× magnification scale. For 20× magnification scale, the sizes are halved

在我们的细分框架,一个小块选择大小为 112×112 像素。我们还探讨了大小对我们的细分框架的影响。结果如表 8 所示。从结果中可以看出，较小的小块尺寸将会产生更好的数据集分割结果。这个事实和我们的预测相同。在分割框架中，对每个采样的训练小块，根据其与带注释的区域的重叠比率，给出阳性或阴性的标签，并根据所有采样的小块的预测标签构造分割结果。在这种情况下，较大的小块尺寸会影响分割区域边界的分辨率，从而影响分割

结果的准确性。

Table 8 Segmentation results of different patch sizes

Dataset	112×112	224×224	336×336
MICCAI brain	84.0%	78.5%	75.7%
CRC	93.2%	86.9%	81.3%

5.4 CNN 激活特征的可视化

我们提出的采用 CNN 特征的框架对脑瘤和结肠癌数据集都有很高的准确性。我们感兴趣的是，我们的分类器从 CNN 的特性中学到了什么，以及它们是否能揭示生物学的真知。为了达到这个目的，最后一个隐藏层(4096 维)的神经元反应的各个组成部分被可视化，以观察 CNN 特征的属性。特别地，我们将他们的形象和特性的反应可视化，以理解我们的 CNN 认为重要的图像的哪一部分。

从图像的角度出发，利用线性支持向量机训练的分类模型，对每个小块进行信任分配。我们将每个小块的置信得分可视化为一个热图(图 5 和图 6)。一个区域越红，分类器就越有信心认为该区域是阳性的(相反，为阴性的)。热图有助于可视化分类器认为的重要区域。对于每个分类任务，每个类别的一个图像都显示在论文中。

在特征方面，我们将在最后一个隐藏层的单个神经元的响应可视化，以观察 CNN 特征的特征(图 7 和图 8)。顶部激活的特征尺寸由分类支持向量机模型的最高权重确定。对于相关的神经元，最能激活它们的小块被选中(在该特性维度中具有最高价值的小块)。

5.5 图像层级的热图

虽然我们没有明确标明每种癌症类型的属性，但我们分类器的热图显示它们确实突出了具有代表性的热点。例如，坏死区域是 GBM 的特征，通常被认为是高度阳性的。

对于脑瘤，热图如图 5 所示。我们将全切片图像分别标记为 GBM 和 LGG。在这个分类场景中，两个类都是胶质瘤，但是有不同的胶质瘤级别。可通过 H&E 染色发现高等级的胶质瘤包括肿瘤星形细胞瘤和多形性胶质母细胞瘤，伴随坏死区域和增生的血管和巨核细胞的出现。在热图的例子中，GBM 的内皮细胞增殖区被很好的捕获。

对于结肠癌，双值和多类分类的热图如图 6 所示。在二元场景中，我们的 CNN 成功地识别了癌症实例中的畸形上皮细胞和正常情况下均匀间隔的细胞结构。例如，在腺癌(AC)亚型的例子中，图中显示的大多数恶性导管分子都是由二元分类器突出显示的。对于其余的图像，基质细胞是丰富的，被认为是中性的或正常的，因为它们在生物学上是良性的。在正常的例子中显示的流明部分被误诊为癌状区域，因为它类似于形状不规则的上皮细胞。然而，每种癌症亚型的某些特征都被二元分类器所忽略。在黏液癌(MC)的例子中，分类器识别致密的上皮细胞，但忽略了 MC 的主要特征，在这里可以看到大量的胞外黏液(原始图像中的浅紫色区域)。这是由于胶体和空区之间的相似性，这使得在二元场景中更难分辨。

在多类场景中，每个子类型的特定特征被强调，在分类器热图中变得明显。在 MC 的例子中，只有胶体部分触发 MC 分类器，其他恶性部分被抑制。他们的分类器成功地捕获了锯齿状癌(SC)和乳头状癌(PC)的独特模式。在 SC 亚型中，不同于所有区域的情况都被认为是恶性的，只有齿状上皮结构仍然高度自信。在 PC 亚型中，只突出了细长的管状结构。分类器忽略了许多独特的 SC 模式，因为它们类似于我们的小块尺度下的 PC 的管状特性。对于筛分型的型腺癌(CCTA)，其独特的筛状特征表现为频繁的穿孔在热图中突出。对于 AC 亚型，许多恶性导管分子被分类器从二分类到多类的场景中忽略，这是由于所有癌症亚型中相似的无所不在的结构，对提高性能没有帮助。对于正常的例子，二分类和多类分类器显示一致的结果，而在多类中，图像中间的流明部分被正确地抑制。

为了比较 CNN 的激活特征和其他特征，手动提取特征的热图也如图 6 所示。从图中我们可以清楚地看到 CNN 激活特性的优势。

5.6 特征块描述

在从不同的医学图像中提取的 CNN 特征中，我们发现单一特征维度可以显示特定的特征，这是在应用 CNN 激活特征的可视化时发现的令人兴奋的发现之一。尽管可能存在某些类型的手动设计的特征提供了相同的特性，但是 CNN 能够自动地从大型图像数据集中学习这些特征，而不需要任何手工设计。由组织病理学家报告，一些特征可以传达临床的见解，这也可以验证我们的发现从图像级的热图分析。每个特征的特征是通过从所有的图像中选择最具权重的图像来可视化的。有关脑瘤图像的更多细节，请参阅[19]。

对于直肠癌而言，两类和多类分类中最具鉴别性的特征是形象化的，如图 7 和图 8 所示。与热图中的发现相似，即使我们没有提供任何病理学特征的额外置信，分类器中的高权重特征对应于一个类别的特定特征。在二分类中，重要的癌症特征包括腺癌(1、2、4、5、6 行)和乳头状(第三行)；虽然重要的非癌症特征包括正常腺体(第一和第二行)，淋巴细胞群(第三排)，出血(第 4 和第 5 行)和脂肪(第 6 行)。

多类分类器自动发现每个子类型更具体的特性，有些情况特别有趣，具有潜在的指导意义。例如，CCTA 的特征不仅包括前面提到的筛状的结构(第二行)，而且还包含了在出血区域(第 1 行)激活的特征，这表明 CCTA 和出血之间存在一些未发现的相关性。

许多 CNN 的特点也为癌症组织分类提供了一些新的标准。例如，PC 特性区分了其特殊管状结构的尖端(第一排)和中段(第二行)。MC 的特征似乎是通过黏液密度来分离胶体分泌的小块:第一排的小块比第二行有更多的黏液。这两个特征在这里看起来非常相似(都显示了结肠导管的致密上皮内壁)，同样可以被识别为腺体结构。在 CCTA 小块的特征中，上述的筛状结构(第 1 行)和出血(第 2 行)的特征都被打开，这都是 CCTA 的典型特征。注意到虽然这里显示的出血小块并不属于结肠癌的性质，但它们仍然可以在最后的 CNN 隐藏层中代表一个神经元，而这通常是由出血的特征触发的。对于正常的类型，包括纵向和横向隐窝的小块(肠腺，第 1 行)或基质细胞的小块(第二行)。

6 结论

本文介绍了用 ImageNet 知识训练的深层卷积激活特征，并将 CNN 模型应用于脑瘤和结肠癌数字组织病理学数据的提取。我们成功地将 ImageNet 知识作为深层卷积激活特性，对组织病理学图像进行分类和分割，训练数据相对较少。根据我们的实验，CNN 得到的特征明显优于手动特征。此外，由于单一组织病理学图像的大小，采用特征池技术在分类框架中构建单一的图像级特征向量。实验表明，我们的框架达到了 97.5%的最先进的分类结果，用于 MICCAI 脑瘤挑战的分割准确率 84%。之后，我们将两个框架应用于结肠癌图像，并取得了相似的成功，比以前的方法有了显著的改进。

此外，我们的分类器所掌握的特征可以具有生物学上的意义，这是病理学家所认可的。联合起来，这些选定的小块或区域的组织病理学形态学将帮助病理学家发现具有生物学洞察力的模式。通过观察 CNN 激活特征神经元的识别斑块，我们可以发现相应亚型的组织成分。研究不同癌症分期和亚型的发展过程是有益的。通过应用数字组织病理学图像分析，可以捕获和量化复杂形态模式的细微差异，并重新研究它们的联合作用，以反映患者的预后或药物反应，并提供细粒度的特征。

我们的动机是为组织病理学问题引入一种通用的解决方案。这使得我们的设置比其他大多数都简单。全卷积网络(FCN)[18]不适合对大规模图像进行分类。因此，我们不将我们的方法与 FCN 进行比较。在今后的工作中，我们将把我们的方法与 FCN 进行分割。