

全数字病理切片的高通量细胞核分割

文可，柴磊，尹智卓，丁峰，王朝晖，吕国锋，安虹

摘要——精准的病理学细胞核分割在各种临床应用和医学研究中起着至关重要的作用。近年来，随着全切片成像技术的成熟，快速且精准的数字化病理数据的分析需求日益增长。本文主要贡献是提出了 P-RCNN，一种基于 Mask-RCNN 的细胞核分割模型。P-RCNN 在 Kaggle DSB 2018 竞赛的测试集上取得 LB-score 0.618（第二名）的成绩；而且泛化能力强，在 TCGA、ISBI2009 和 TNBC 上，均可达到当前最先进的分割分数。我们还从文件 I/O 和参数服务器分布式训练两个方面加速 P-RCNN 的训练过程，并通过把模型移植到 C#来加速预测。

关键词——病理学；细胞核分割；P-RCNN；Mask-RCNN；参数服务器；

High Throughput Nuclei Segmentation for Pathology Whole Slide Images

ABSTRACT - Precise pathological nuclei segmentation is of vital importance to a variety of clinical applications and academic researches. Recently, with the development of Whole Slide Imaging technology, the demand of analyzing digital slides rapidly and precisely grows stronger. The major contribution of our work is to create P-RCNN, which is a nuclei segmentation model based on Mask-RCNN. P-RCNN win an LB-score of 0.618 on the test set of Kaggle DSB 2018 competition, which could be the second place. Moreover, it can generalize to many other datasets such as TCGA, ISBI2009 and TNBC, and get state-of-art segmentation scores. In addition, we accelerate the training process in terms of file I/O and distributed training with Parameter Server, and accelerate the inference process by porting the model into C#.

Key Words - Pathology; Nuclei Segmentation; P-RCNN; Mask-RCNN; Parameter Server;

1. 引言

1.1. 研究动机

病理诊断是各种疾病诊断的“金标准”。病理的诊断模式是把组织细胞形态与临床、影像和检验结合起来进行综合分析和判断，对疾病做出最终的诊断。病理组织细胞形态的分析是采用 H&E 染色、IHC（免疫组化）、分子病理等手段检查给定组织的细胞形态特征，是病理诊断技术的核心。

精确的病理组织细胞形态检测在各种临床应用和医学研究中起着至关重要的作用。例如，病理学家通过在活组织切片上检查给定组织的细胞形态和 IHC 特征，可以获得对肾透明细胞癌和前列腺癌扩散因子 MIB-1 表达的评估，进而诊断出肿瘤的良恶性和分级^[1]。传统上，病理学家通过光学显微镜用肉眼进行组织、细胞形态特征的分析 and 评估。人工病理分析存在以下缺陷：

- (1) 由于病理图像尺寸巨大，组织类型多，大多为良性和正常组织，分析十分耗时；
- (2) 无法使用定量的图像特征和度量方法；

(3) 分析的结论会随专家和单位的不同而不同，不同专家之间的一致性非常低。

病理诊断方法在良恶性鉴别诊断方面缺乏客观标准，对恶性肿瘤进行分类有时十分困难，无法满足临床需要。

利用数字图像处理、计算机视觉和深度学习方法，自动化分析病理组织细胞的异型性（形态、染色等特征），是病理分析和诊断技术的新兴研究领域。近年来，许多病理实验室开始采用全切片成像技术，用高分辨率、高通量切片扫描仪将组织切片数字化，在为定量评估细胞的形态学特征提供了巨大机遇的同时，也使得病理诊断的瓶颈从数据采集分析的数量和速度，转变为数据分析处理的精度和效率。为了辅助病理学家分析海量的数字化病理图像，我们需要一套基于数字切片辅助病理诊断的自动化工具，旨在：

(1) 提高诊断质量。自动化分析可以提高病理组织细胞的形态学特征分析的量化水平，使得分析结果具有可重复性，更加客观精确；

(2) 提高诊断效率。自动化分析可以利用高性能计算技术实现对海量病理数据的高通量分析，具有人工无可比拟的诊断效率；

(3) 提升病理学家的工作体验。传统上，使用光学显微镜进行组织细胞形态学特征分析需要日复一日、年复一年的重复性工作，单调乏味。自动化分析可以把病理学家从繁重的劳务性工作中解放出来，更多地去发掘细胞的形态学特征。

细胞核分割的挑战在病理学图像处理领域中，细胞核分割是分析细胞形态等特征的第一步，也是最难的一步。在全数字切片上进行细胞核分割面临四个挑战：

(1) 病理切片数据量大。用于计算机视觉图像识别的大型可视化数据集 ImageNet^[2] (1.4×10^8 张手动注释图像)，仅相当于病理组织中 1136 张切片的数据量，而这尚且不足中国科学技术大学附属第一医院（安徽省肿瘤医院）每三天新增的切片数量；

(2) 专业的病理学知识需求。和计算机视觉常见的场景相比，为细胞核分割设计计算机视觉方法之前，必须

通晓细胞病理学的知识。各种恶性肿瘤的诊断中，仅依靠细胞形态学上的特征无法得到正确的判断；

(3) 需要精确、高效的分割算法的支撑。和自然场景分割任务（ImageNet dataset: 1.04 object/image; COCO^[3] dataset: 7.18 object/image）不同的是，病理图像中细胞核数量多、尺寸小、分布密集（细胞核数据集 TCGA^[4]: 565.53 object/image）。这要求分割算法不仅能分割出细胞核，更要精确地区分开每一个细胞核；

(4) 需要高性能计算的支撑。目前 ImageNet 训练时间记录是 6.6 分钟^[5]，它使用了 2048 块 NVIDIA Tesla P40 加速卡，而一家肿瘤医院每三天病理组织每日新增的切片数据量就相当于 1 个 ImageNet 训练的数据量。要想实现自动化的细胞核分割，必须要有高性能计算的支撑，才能在可接受的时间内得到分析结果。

1.2. 本文的主要贡献

本文提出了 P-RCNN 模型，用来对病理组织全数字切片的进行高通量的细胞核分割。该模型适用性广泛，可从病理学常见的 H&E 染色、IHC、分子病理等扫描切片中精确地分割出细胞核的轮廓。

我们从三个方面提高了模型的精度：

(1) 面临数据质量差，有标记病理数据少的问题，提出一种病理图像数据预处理方案，有效地提高数据多样性，增强模型的泛化能力；

(2) 对通用的分割模型——Mask RCNN 进行调优，以应对病理图像细胞核数量多、尺寸小、分布密集的特点，显著地提升细胞核分割的效果；

(3) 提出了一种分块预测及合并的方案，有效地提高对大的病理图像的预测准确性，这也是全数字切片预测的基础。

我们从三个方面提高了模型训练和预测的速度：

(1) 优化文件 I/O，提出一种简单的数据压缩方案来缓解病理图像占用内存大、I/O 次数多的问题，显著减少在线学习的训练时间；

(2) 使用 Parameter Server 进行分布式训练，从而成倍

地加快模型的收敛速度；并使用 InfiniBand+RDMA 技术来降低通信延迟；

(3) 将模型移植到 C#上，提高预测速度。

2. 研究现状

2.1. 细胞核分割的发展历史

细胞核分割的研究可以追溯到 20 世纪 90 年代。在过去的几十年中，计算能力、图像扫描技术、图像处理和识别算法的进步促进了病理学图像分析的发展。然而，算法输出和病理专家的注释之间仍然存在明显的差距。传统的细胞核分割技术有：

(1) 基于强度阈值的 Otsu 算法^[6,7]，通过设定不同的特征阈值，对像素点进行分类；

(2) 基于拓扑理论的 Watershed^[8]算法，根据分水岭的构造来考虑图像的分割；还有基于图论，区域和遗传算法等的分割算法^[9]。

近年来，基于深度学习的图像分割方法变得越来越有吸引力。神经网络有着强大特征表示能力，这极大地促进了的图像识别相关任务的进步。基于深度学习的分割算法主要有两种思想：

(1) 基于全卷积神经网络的 FCN^[10]，U-Net^[11]算法，通

过卷积与反卷积，对图像的每一像素作二分类（细胞核或背景）；

(2) 基于候选区域和分类思想的 R-CNN 算法^[12]，通过选择性搜索（或基于深度神经网络的 RPN 模型^[13]）提取候选区域并对其作分类与分割。

所有上述的分割方法都有一个共同的缺点：它们仅对一种或几种特定类型的病理组织切片有效，并且通常对手动设置的参数高度敏感。然而病理组织切片来源广泛，很难开发出一种泛化能力强适用于不同病理组织切片的单一模型或方法。如图 1 所示，在 H&E 染色图片上，Otsu 算法存在“欠分割”现象（即每个分割区域包含多个相邻的细胞核对象），通俗地讲就是分割得不够细致；Watershed 算法存在“过分割”现象（即每个分割区域仅包含一个细胞核，而且仅占该细胞核的一小部分区域），通俗地讲就是分割得太细碎。

在 Mask-RCNN 出现之前，U-Net 常常被用来做细胞核分割。U-Net 的分割结果虽然能覆盖图像中大部分细胞核，但难以分开粘连的细胞核现象。我们基于 Mask-RCNN 提出了 P-RCNN 分割模型，它泛化能力强，针对不同染色病理组织切片（H&E 染色，荧光图像及低倍镜细胞核）都能达到当前最好的分割效果。这是因为 Mask-RCNN^[14]可以避免欠分割和过分割的现象，还可以区分开粘连的细胞核。这对分布密集的路径组织（如癌变的区域细胞核分布往往十分密集）切片尤为重要。

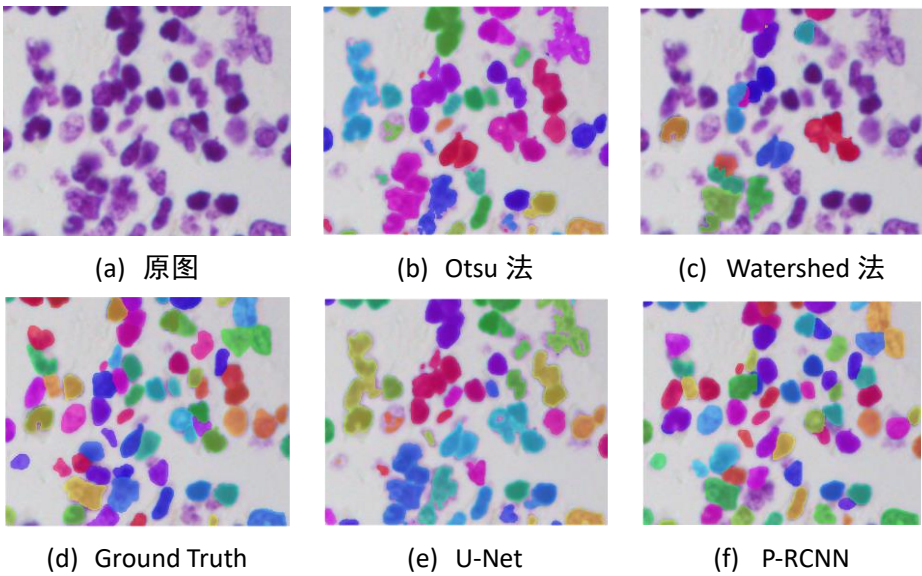


图 1 P-RCNN 和较早的分割算法 Otsu、Watershed、U-Net 在细胞核分割任务中的效果对比

2.2. 公开的病理数据集

图像分割是生物医学图像分析的基础，近年来，基于深度学习的方法在许多生物医学图像分割基准测试上都取得较好的成果。基于深度学习的方法，因为其精确度高，泛化能力强的特性，已经逐渐成为了生物医学图像分析的主流选择。尽管在生物医学图像分析领域上，深度学习的算法已经取得巨大的成功，但是仍然面临一个关键的障碍：由于基于深度学习的自然场景应用通常需要大量高精度标注的训练数据，而区别于计算机视觉的自然场景，存在两个主要原因，在智能医疗

领域上获得足量的训练数据十分困难：

- (1) 只有经过专业训练的生物医学专家才能对数据进行高精度标注，无法利用人群对海量的数据作标注；
- (2) 生物医学领域的图像通常比自然场景的图像有更多更小的实例，如一个 1000×1000 的细胞组织切片可能有多达 1000 个细胞，人为手动的数据标注不现实；

本文介绍的细胞核分割应用中，我们主要引用以下四组公开的病理组织切片数据集，如表 1，图 2 所示，使用 K-Means 算法^[15]将病理组织切片数据集分为三类（H&E 染色，荧光图像及低倍镜细胞核）：

数据集	平均尺寸（像素）		图片数	细胞核数	细胞核数下界	细胞核数上界	H&E染色	荧光	低倍光镜
	高	宽							
DSB_train	333.9	378.5	664	21547	1	375	107	541	16
DSB_Test	382.5	430.9	3019	-	-	-	219	2497	303
ISBI1	1030.0	1349.0	48	1831	24	63	-	48	-
ISBI2	1024.0	1344.0	49	2178	29	70	-	48	-
TNBC	512.0	512.0	50	4028	3	297	50	-	-
TCGA	1000.0	1000.0	30	21541	294	557	30	-	-

表 1 四组数据集（DSB 2018, ISBI, TNBC, TCGA）统计信息

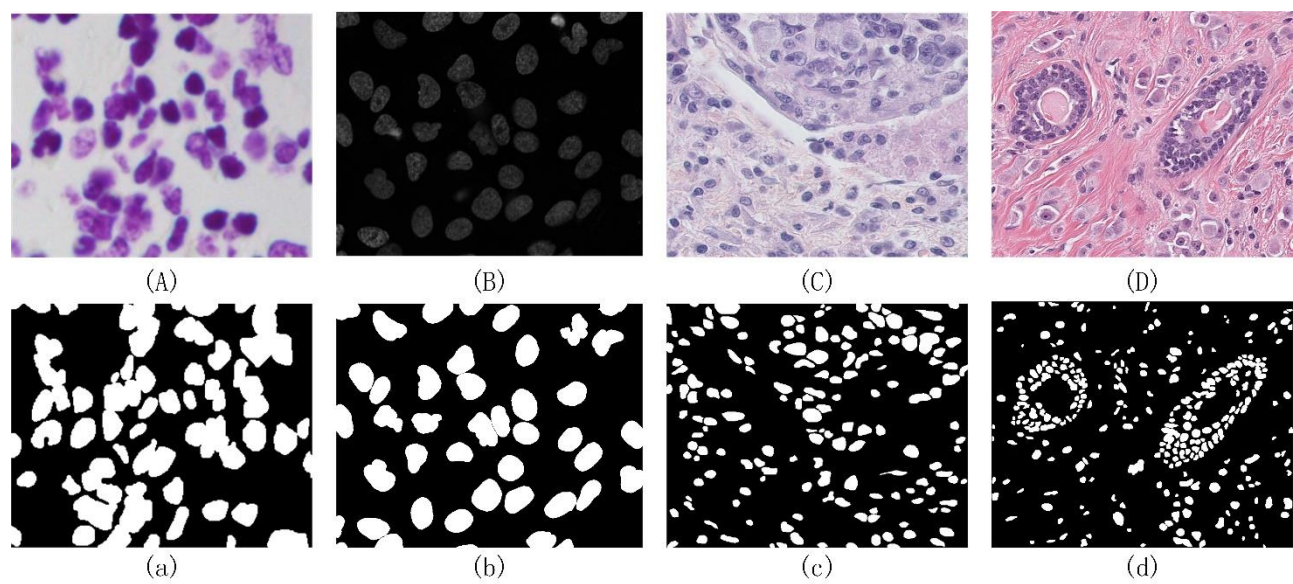


图 2 四组数据集，(A) DSB, (B) ISBI2009, (C) TNBC, (D) TCGA

DSB: 该数据集来自 Kaggle 于 2018 年举行的数据科学碗 (DSB) 比赛, 标注信息由哈佛大学和麻省理工学院的 Carpenter 实验室提供。DSB 训练集中主要包含三类的病理组织切片图像, 共 664 张。一共有 21,547 个细胞核, 细胞核分布不同, 单张图像最少/最多包含细胞核差距大 (1/375)。测试集未公布细胞核标注信息, 共有 3019 张图像, 可见 DSB 训练与测试集的数据分布差距大, 对算法泛化能力要求高;

ISBI2009: 该数据集来自^[16], 标注信息是由卡耐基梅隆大学的 Murphy 实验室提供。数据集分为两部分, 第一部分有 49 张 1030×1349 图像, 均采集自人体骨肉瘤细胞 (U2OS cells), 包含 1,831 个细胞核; 第二部分有 48 张 1024×1344 图像, 均采集自小鼠胚胎成纤维细胞 (NIH3T3 cells), 包含 2178 个细胞核; 数据均属于第二类图像, 细胞核分布平均;

TNBC: 该数据集来自^[17], 标注信息是由巴黎文理研究大学的 CBO 实验室提供。该数据集有 50 张 512×512 的图像, 采集自三阴性乳腺癌细胞, 一共有 4028 个细胞核, 均属于第一类图像。数据集中细胞核分布不同, 单张图像最少/最多包含细胞核 (3/297);

TCGA: 该数据集来自^[4], 标注信息是由印度理工学院电子与电气工程系提供。该数据集有 30 张 1000×1000 的图像, 分别采集自七种不同器官 (胸腺, 肝, 肾, 前列腺, 膀胱, 结肠和胃) 的 40 倍 H&E 染色切片, 一共有 16966 个细胞核, 均属于第一类图像。数据集中细胞核分布密集, 差距大, 单张图像最少/最多包含细胞核 (294/557)。

2.3. 分割精度的评估标准

在计算机视觉领域, 通常使用以下两类评估标准 (像素级和对象级) 来衡量分割算法的分割精度。

2.3.1. 像素级的评估标准

这是最直观的分割评估标准。设 n_{ij} 是第 i 类像素被预测为第 j 类像素, n_{class} 表示为不同类别的个数, 考虑我们的任务是细胞核分割, 即 $n_{class} = 2$, 令 $t_i = \sum_j n_{ij}$, 即所有被预测为第 i 类的像素数。以下三种标准是语义分割中最常用的评估标准^[10]:

$$(1) \text{ Pixel Accuracy: } \frac{\sum_i n_{ii}}{\sum_i t_i}$$

$$(2) \text{ Mean Accuracy: } \frac{1}{n_{class}} \cdot \sum_i \frac{n_{ii}}{t_i}$$

$$(3) \text{ Mean IU: } \frac{1}{n_{class}} \cdot \sum_i \frac{n_{ii}}{t_i + \sum_j n_{ji} - n_{ii}}$$

像素级的评估标准只能在一定程度上反映分割算法的性能, 因为它对细胞核间距不敏感, 惩罚力度小, 即使是“欠分割”的现象, 也能在得到很高分数。考虑到要对病理组织数字切片的细胞核做形态学量化分析, 需要精确的划分开每一个细胞核, 这需要更加精准的评估方案。

2.3.2. 对象级的评估标准

主要基于连通域划分对象。我们使用 IoU (也称 Jaccard index)^[18] 判断预测对象是否“命中”:

$$\text{IoU}_{ij} = \frac{|G_i \cap S_j|}{|G_i \cup S_j|}$$

G_i 表示第 i 个细胞核真实位置, S_j 表示第 j 个预测细胞核位置。通过设定一定的阈值 t (threshold) 获得:

- (1) True positive (TP), 正确预测为细胞核的个数;
- (2) False positive (FP), 错误预测为细胞核的个数;
- (3) False negative (FN), 未被预测为细胞核的个数;

我们使用以下三种对象级评估标准, 它们是实例分割和目标检测中常用的标准:

(1) F1-score^[19]:

$$\frac{2 \times TP}{2 \times TP + FN + FP} \cdot \text{这里我们取阈值 } t = 0.5.$$

(2) Aggregated Jaccard Index (AJI)^[4]:

$$\frac{\sum_i G_i \cap S_j}{\sum_i (G_i \cup S_j) + \sum_{k \notin J} |S_k|}$$

每个 $j = \arg \max_j \left(\frac{|G_i \cap S_j|}{|G_i \cup S_j|} \right)$, 其中 $t = 0.5$ 。

(3) Average Precision (AP)^[3]:

$$\frac{1}{|T|} \sum_{t \in T} \frac{TP(t)}{TP(t) + FP(t) + FN(t)}$$

其中 $T = \{0.5, 0.55, \dots, 0.95\}$ 。

对象级的评估方案往往更能真实反映算法的分割性能, 其中 AP 是最重要的一个评估标准, 在 COCO 和 Google Open Images Challenge 中, 均使用 AP 作为最终评分标准^[20]。

3. 方法与创新

3.1. 数字病理切片的预处理和后处理

深度学习领域中, 数据处理是深度学习中极为重要的一部分, 可以说输入数据的好坏决定了这一任务的性能上限, 而算法和模型做的就是逼近这一上限。区别于计算机视觉的自然场景, 病理组织切片数据获取途径的独特性, 决定了病理组织切片数据的复杂性。因此好的病理数据处理方案在生物医学图像分析领域显得尤其重要。本文中, 我们使用以下三种数据处理的方案:

3.1.1. 在线数据增强

深度学习模型往往有上千万的参数 (本文使用的分割算法参数量: 6.3×10^7) 需要学习拟合, 大规模的数据集可以避免过度拟合问题, 由于病理组织切片数

据收集困难 (ImageNet: 128 万张图片 1000 类; DSB: 664 张)。在缺乏大量数据的前提下, 数据增强显得尤为重要, 我们使用在线数据增强 (区别离线, 减少数据存储空间的开销, 最大限度的增加数据多样性), 对即将输入模型的小批量数据以一定概率 (概率设定为 0.5) 执行相应的变化:

- (1) 透视变换, 将图片投影到另一个新的视平面;
- (2) 裁剪, 随机从原图中裁剪出一定大小 (256×256) 的图像;
- (3) 图像的水平翻转、竖直翻转、 90° 旋转。

3.1.2. 颜色正则化

对 H&E 染色图像来说, 由于染色剂、染色过程和扫描仪的不同, 都会导致图像颜色的巨大差异。我们使用颜色正则化^[21], 通过基于稀疏非负矩阵分解 (SNMF) 的颜色混合建模可以在保留生物结构的前提下, 改善图像颜色, 达到提升算法分割性能的效果。该技术需要专业病理医生指定一张认为染色效果最好的目标图像, 通过计算其染色密度图, 计算其他图像得到颜色正则化后图像, 如图 3 所示

3.1.3. 数据分块及随机过采样

从表 1 可发现, 病理组织切片的数据大小不一, 在 Mask R-CNN 的训练过程中, 我们需要将图像统一调整至一个固定大小。但考虑到病理图像的细胞核特性 (数量多, 尺寸小, 分布密集), 大图像调整至小图像会损失一定细胞核形态特征信息甚至细胞核变得太小而成为无效信息, 我们使用该调整方案^[22], 保持原图长宽比对图像作调整, 保证调整后图像信息损失量最少。而且数据类型分布不平衡, 不同类型病理组织切片数目差别很大, 我们所使用的数据集中, 大量图像属于第二类图像。故我们使用数据分块及随机过采样方案, 即对第一, 三类图像作“过采样”, 平衡三种不同类型病理切片数据。不再采取调整方案, 对小图像采用随机采样的方案, 而对大图像则先作分块, 再从每一分块中随机采样作为输入数据。

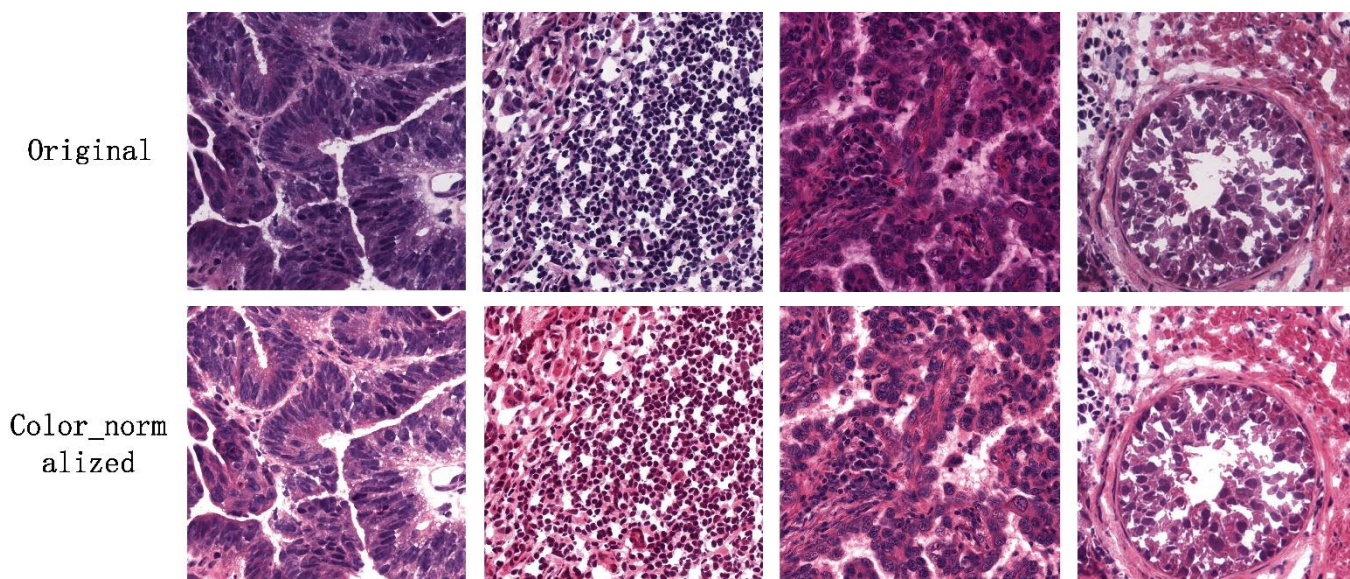


图 3 颜色正则化前后对比

3.2. 模型设计

本文提出了基于 Mask R-CNN 的 P-RCNN 模型。后者主要分为 RPN 模型（候选区域提取）和分类模型（候选区域分类及 Mask 生成）。

为了适应细胞核分割以应用，我们在 Mask R-CNN 模型的基础上作出以下的精细调优：为了保证输入数据的完整性，采用^[22]，将图像调整至 512×512 ；底层神经网络参数的初始化使用 ImageNet 数据集中预训练好的 ResNet101 的参数；在共享特征图的生成中，我们利用 Feature Pyramid Net (FPN) 的多尺度特征融合思想^[22]，底层的特征语义信息比较少，但目标位置明确；高层的特征语义信息比较多，但目标位置比较粗略。从 ResNet 网络中，提取其中每一阶段最后一层共 4 层作为中间层通过 FPN 生成 5 个共享特征图；基于病理组织切片中，我们考虑细胞核数量多，尺寸小，分布密集，故提高 5 个共享特征图的 Anchors 尺寸至 [8,16,32,64,128]；提高训练过程的有效候选区域数至 512 个，及每一图像最大预测对象数至 400 个；其余网络参数设定和原文 Mask R-CNN 相同。使用 Adam 梯度下降优化算法^[23]，共训练 75 轮(epoch)，学习率每 25 轮递减 0.1 倍，初始学习率为 0.0001。

3.3. 分块预测与合并

本文所用的训练数据只有长宽不足 2000 像素的小尺寸图像，而全数字切片分辨率高 ($40,000 \times 60,000$)，不仅标注成本极高，单张像素量就超过我们所使用的数据集的总和，几乎不可能将单张切片存入计算机作训练及预测。因此，对全片作分块处理必不可少。对分块进行预测很简单，但合并起来有难度，需要保证位于边缘的、被切割的细胞核能够重新拼合成“平滑”的细胞核。针对这个问题，我们提出了这种分块预测及合并算法：

- (1) 找到每一块图片的边缘细胞核；
- (2) 对每一个边缘细胞核，找到另一侧与它拼合长度最大的细胞核，我们认为这两个细胞核属于同一个细胞核；
- (3) 从两张图的边缘细胞核集合删去被合并的细胞核，重复步骤 (2)，直至其中一张图的边缘细胞核集为空或并集最大值小于一定阈值。效果如图 4 所示。

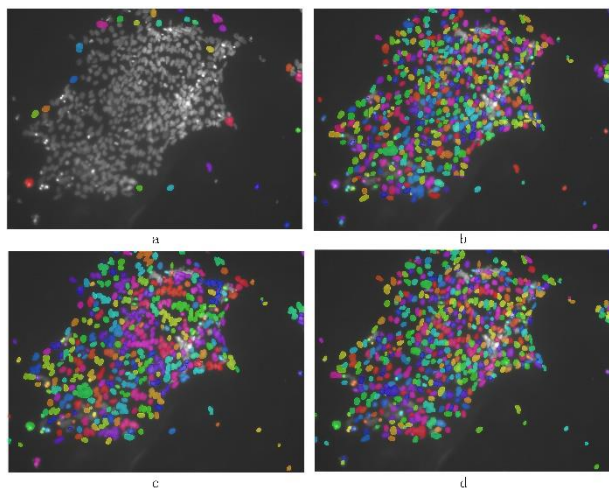


图 4 大图像的分块预测与合并效果, a. 直接缩放至小图像预测; b. 简单分块预测, 未合并; c. 简单分块预测, 合并所有邻接连通域; d. 我们提出的分块预测及合并算法

4. 优化方案

我们的测试平台是一些 CPU 节点, 每个节点有 2 块 CPU, 型号为 Intel(R) Xeon(R) CPU E5-2695 v4。每个节点有 TensorFlow-1.10.1 的 MKL 版本及其 GPU 版本。我们在这些平台上进行了文件 I/O 优化、基于参数服务器的分布式训练加速, 以及预测加速。

4.1. 文件 I/O 优化

为了适应系统设计里对在线训练的要求, 我们在给 P-RCNN 输入训练数据时, 没有把整个训练集读入内存 (这在全片应用中也不现实), 而是按需读取。这就使得 I/O 延迟散布在训练过程中每一步迭代里。因此我们需要对文件 I/O 进行优化, 提高每一步迭代的速度。我们从以下两个方面优化了文件 I/O, 并显著加快了训练速度:

4.1.1. 使用 HDF5 格式

我们把训练图片事先转换成 HDF5 格式^[24], 利用 HDF5 格式能直接读取为 NumPy 数组的特性来加快图

片读取。我们在 TCGA 数据集上测试了加速效果, 实验发现每轮训练的速度和读取原始图片相比提高了 9.2 倍 (如图 5)。格式转化与数据种类无关, 所以我们有理由认为, 这种转化在其他训练数据上一样可以带来显著的加速。

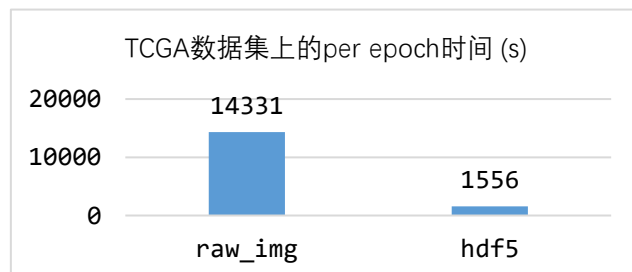


图 5 在 TCGA 数据集上采用 HDF5 格式转化, 使得每轮训练加速了 9.2 倍

需要注意的是, 把原始图片转化为 HDF5 格式的开销并不需要考虑。这部分开销被转嫁给医生客户端, 因为在我们的应用设计中, 格式转化由医生客户端来做。

4.1.2. 数据压缩和打包

我们进一步对训练数据进行压缩和打包。一种直观的压缩方法是切除每个 Mask 的黑边。考虑到 Mask 是二值图像, 有很多连续、重复的像素, 我们对 Mask 进行 GZip 压缩。为了减少文件 I/O 次数, 我们还把一张图片和它所有的 Masks 打包进同一个 HDF5 文件中。在 TCGA 数据集上, 这些手段使得每轮训练的速度进一步加速了 1.6 倍 (如)。数据压缩和打包也是与数据种类无关的, 可以推广到任意数据集上。

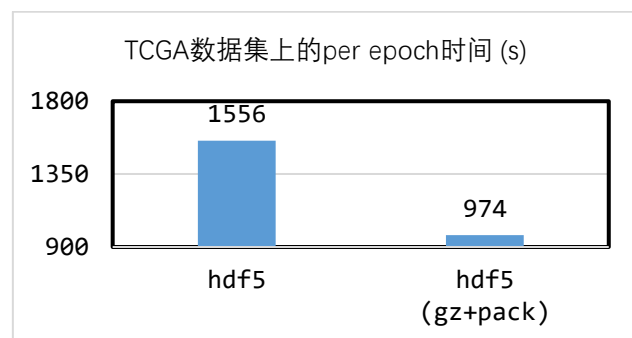


图 6 在 TCGA 数据集上采用数据压缩和打包, 使得每轮训练进一步获得 1.6 倍的加速

类似于 4.1.1，压缩和打包的开销也可以转嫁给客户端。

4.2. 基于 Parameter Server 的分布式训练加速

P-RCNN 有 6.3×10^7 个可训练参数。这种参数量在单机上需要训练 8.7 小时才能使损失函数收敛到 1 以内。

参数服务器 (Parameter Server) 通过让每个 Worker 只计算一部分参数的梯度来实现异步的分布式训练加速^[25]。我们采用 1 Server, N Worker 的集群架构来训练模型，从同样的初始权重开始训练，观察损失函数 <1 所需的时间。我们以 Loss-Time 曲线的形式来展示分布式训练加速的效果，如图 7 所示。1、2、4、8 个节点上分别需要 31269、25823、21009、16697 秒使得损失函数收敛到 1 以内。

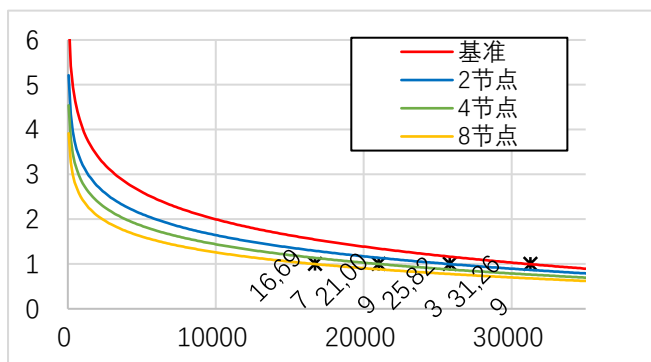


图 7 模型在 1、2、4、8 个节点上进行分布式训练的 Loss-Time 曲线变化趋势

之所以选择损失函数 <1 为观测标准，是因为 P-RCNN 的模型训练到损失函数小于 1 时，模型在验证集和测试集上的表现就已经可以接受。

4.3. 迁移到 C# 的预测加速

为了提高每个小片的预测速度，我们还尝试了把模型从 TensorFlow Python 框架迁移到 TensorFlow C# 框架上，获得了 14 倍的加速，如图 xxx。

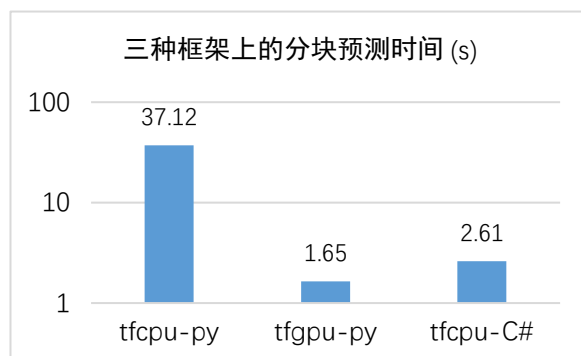


图 8 将模型移植到 TensorFlow C# 获得了接近于 GPU 预测的速度。

注：tfgpu-py 所用的 GPU 为 NVIDIA Tesla V100。

5. 实验结果与分析

5.1. 细胞核分割的精度

我们将四组数据集 DSB, TCGA, TNBC, ISBI2009 合并成一组大的数据集 D (共 1637 张图像) 用来训练和测试 P-RCNN。我们对 D 中的数据作过采样操作。按照 Kmean 算法分类比例选择其中 10% 数据 (共 164 张图像) 作为验证集。我们观测 2.3.1 和 2.3.2 节提到的六种打分标准，重点关注 2.3.2 节的三种打分标准 (F1-score, AII-score, AP)。

我们分别在 DSB 2018、TCGA、TNB、ISBI 四组数据集上把 P-RCNN 和当前最先进的分割算法^[4, 16, 17, 26-28]作对比，实验结果表明 P-RCNN 的分割效果更加优秀，泛化能力强。P-RCNN 在后三组数据集上表现出当前最好的分割效果。

5.1.1. DSB 2018 测试集上的精度

我们的 P-RCNN 在 DSB 2018 测试平台上得到了 $AP=0.618$ 的分数，该分数在 LeaderBoard 上能排到第 2 名 (已公开的第一名 $AP=0.631$ ，第二名 $AP=0.614$)。此外，我们将结果与 U-Net (2-class, 3-class)^[27]，StarDist^[27]及 Mask R-CNN^[14]作对比，在 AP (0.5:0.95) 标准上比 Mask R-CNN 有明显的提升 (+0.0674)；在更严格的 DSB AP 标准上，我们的分数仍然能达到 0.618，而且与 Mask R-CNN 差距更大 (+0.134)，这表明我们

的细胞核分割算法的泛化能力强。实验结果如表 2 所示。

模型	AP	DSB Test AP
U-Net(2-class)	0.4916	0.3517
U-Net(3-class)	0.5942	-
Mask R-CNN	0.6288	0.484
StarDist	0.5984	-
P-RCNN	0.6962	0.618

表 2 DSB 数据集不同算法分割效果对比，其中 AP 计算阈值 {0.5,0.55,...,0.90}，DSB AP 的阈值范围为 {0.5,0.55,...,0.95}，所以 DSB AP 比 AP 更严格

5.1.2. TCGA 验证集上的精度

TCGA 数据集只有 30 张来自七种不同器官的 H&E 图像，因此需要做数据扩增。我们对数据作分块随机采样处理：将每张图像分为 4 块大小为 500 × 500 的图像，并且从每一块随机截取 5 张大小为 256 × 256 的图像，从而扩增出 600 张图像。和 CNN2^[4]，CNN3^[4]及 NB Model 相比，P-RCNN 在此数据集上取得了当前最好分割效果（如表 3）：

- (1) F1 分数比 NB Model^[26]高出 0.0285，比 CNN3 高出 0.0558；
- (2) 用更严格的 AJI 标准来比，P-RCNN 比 CNN3 有明显的提升 (+0.3605)
- (3) 需要注意的是，P-RCNN 的 AP-score 为 0.4842，是四组数据集中 AP-score 最低的。我们提出的解释是：

H&E 染色切片中细胞核和背景的差异不如荧光切片明显，而该数据集中全都是 H&E 切片，所以 AP-score 最低。此外，我们没有找到其他模型在此数据集上的 AP-score。

模型	F1-Score	AJI-Score	AP
CNN2	0.7538	0.3482	-
CNN3	0.8267	0.5083	-
NB Model	0.8540	-	-
P-RCNN	0.8825	0.7087	0.4842

表 3 TCGA 数据集上的 F1-Score、AJI-Score、AP-Score 的对比

5.1.3. TNBC 验证集上的精度

TNBC 数据集有 50 张大小为 512 × 512 的图像，由于数据太少，需要进行数据扩增。类似于 5.1.2 节中对 TCGA 数据集的处理方法，我们把数据集扩增为 250 张 256 × 256 的图片。和 PangNet，DeconvNet，FCN 及 Ensemble 四种算法相比，P-RCNN 在此数据集上可达到当前最好的分割效果。从表可以看到在 F1-Score 上比上述四种算法都有明显的提升，至少提升 (+0.0964)，且在 AP 可达 0.5636。P-RCNN 模型的 Pixel Accuracy 评分可达到 0.9547。虽然比其他算法高出的分数不多，但我们认为 Pixel Accuracy 是一种像素级评估，无法精确地反映算法间的精度差距。

模型	Pixel acc.	Mean IU.	F1-Score	AP
PangNet	0.9240	0.7220	0.6760	-
DeconvNet	0.9540	0.8140	0.7820	-
FCN	0.9440	0.7820	0.7630	-
Ensemble	0.9440	0.8040	0.8020	-
P-RCNN	0.9547	0.8442	0.8984	0.5636

表 4 TNBC 数据集上的 F1-Score、AJI-Score、AP-Score

5.1.4. ISBI2009 验证集上的精度

ISBI2009 数据集共有 97 张图像大尺寸图像，但都是荧光切片。考虑到这一点，我们仅对数据作分块处理，不采用随机采样：每张图像作四分块，共 388 张图像。和 Otsu，Watershed，Merge Algo 和 ITMC 四种算法 [16][29]相比，P-RCNN 在此数据集上达到当前最好的分割效果，如所示。此外，在此数据集上的 AP-score 是四种数据集里最高的评分 (0.7167)。考虑 ISBI2009 数据集均是荧光切片，我们认为，算法在荧光切片上可达到最好的分割效果，可以用于对细胞核的形态学量化分析。

模型	pixel acc.	F1-Score	AP
Otsu	0.8300	-	-
Watershed	0.8450	-	-
Merge Algo	0.8950	-	-
ITMC	0.9268	0.8350	-
P-RCNN	0.9733	0.9229	0.7167

表 5 ISBI2009 数据集上的 Pixel Accuracy、F1-Score 和 AP-Score

5.2. 模型调优的作用分析

5.2.1. 对单独的数据集进行微调训练

我们以 5.1 节的模型为基础,分别在 TCGA、TNBC、ISBI2009 训练集上进行微调训练。共训练 45 轮,学习率每 15 轮递减 0.1 倍,初始学习率为 0.0001。下表是我们在针对数据集作微调前后的分割效果对比,由表可看到,微调后,算法在不同数据集上都有一定性能提升。但是可以注意到,对于 TCGA、TNBC 的 H&E 染色数据类型,性能提升不明显,我们认为算法在这一数据类型下学习能力欠佳,无法进一步学习到更多的 H&E 特征信息,故即使是微调后,也无法有明显的性能提升。但是从 ISBI2009 数据集中,微调后在 AP 分数上有明显的提升 (+0.1545),我们认为算法在荧光数据类型下学习能力较强,在只针对该数据类型作微调后,算法可以学习到更多荧光数据的特征信息,故有明显的性能提升。效果如表 6 所示。

数据集	方法	F1-Score	AP
TCGA	P-RCNN	0.8825	0.4842
	Fine-Tune	0.8884	0.4944
TNBC	P-RCNN	0.8984	0.5636
	Fine-Tune	0.8833	0.5727
ISBI2009	P-RCNN	0.9229	0.7167
	Fine-Tune	0.9555	0.8712

表 6 是否使用微调训练的分割精度对比

5.2.2. 在线数据增强

表 7 是我们在使用在线数据增强前后的精度数据。

从表中我们可以看到使用在线的数据增强能明显的提高 AP 的分数,而且在数据更为广泛的 DSB Test 集上,AP 提升更加明显,说明使用在线数据增强方案后,算法的泛化能力更强。而且从模型的收敛性曲线(图 9)也可以明显看到:当未使用在线数据增强,算法模型在训练集上性能表现极强,算法收敛值更低,但是验证集上性能波动特别大,后期甚至呈反弹趋势。显然未使用在线数据增强算法模型存在过拟合的现象,我们认为以一定概率对数据进行变换的在线数据增强方案,能增加了数据多样性,有助于提升算法模型的泛化能力。

方法	F1-score	AP	DSB AP
无增强、无分块合并	0.9217	0.5925	0.519
无增强、有分块合并	0.9204	0.5901	0.534
有增强、无分块合并	0.9276	0.6422	0.612
有增强、有分块合并	0.9121	0.6332	0.618

表 7 是否使用在线数据增强的精度数据



图 9 使用在线数据增强前后的收敛性差别

5.2.3. 分块预测及合并

考虑到分块预测我们针对的是大图,由上图可看到,分块预测及合并方案在分割效果上的提升十分有限,这是因为验证集中只有少量的大尺寸图像,因此我们将所有大图像提取出来,单独验证该方案的性能提升。这是十分有必要的,因为病理临床面临的都是尺寸十分巨大的全片。我们从 DSB 2018 训练集中选取尺寸

大于 1000px 的图像（共 23 张）作为验证集，验证该方案的性能提升。从图 10 可以看到，未使用分块预测及合并方案的预测结果有两种情况：

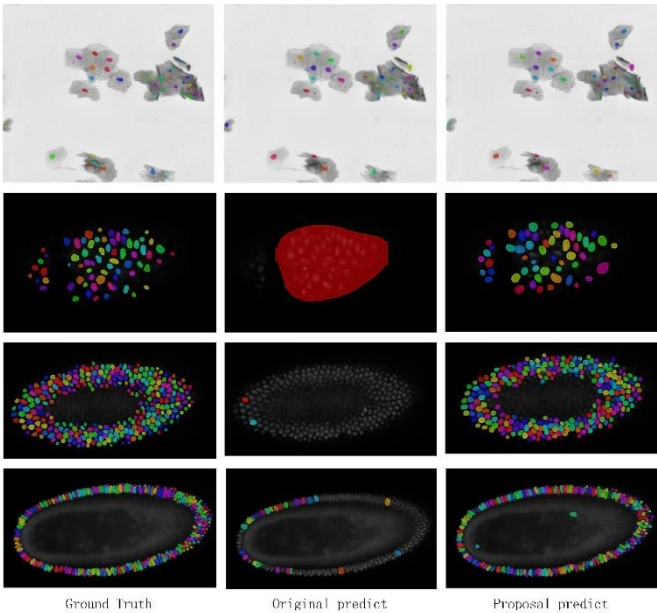


图 10 是否使用分块预测及合并的效果对比图

- (1) 结果不够好（如图 10 行 1）无法预测出全部细胞核，AP 较低；
- (2) 更普遍的情况是几乎无法预测出任何细胞核（如图 10 行 2~4），因为大尺寸图像在缩放后损失大量细胞核信息。由实验结果可见，使用分块预测及合并方案，可以大幅度的提升预测结果（+0.1284）。

image_id	尺寸	AP	AP （分块预测 及合并）
1	[1024, 1024]	0.4118	0.5995
2	[603, 1272]	0.0000	0.1998
3	[603, 1272]	0.0035	0.3362
4	[1024, 1024]	0.0194	0.3871
平均	-	0.3091	0.4375

表 8 分块预测及合并对比实验，仅针对大尺寸图像（23 张），随机挑选四张典型图像作细节分析

6. 总结与展望

细胞核分割算法的高效实现，可以提高大规模组织病理学图像分析的效率，但是仍然存在着很大的发展空间。我们仍将继续讨论研究下一步的优化方向：针对细胞核边缘轮廓预测不够精细，考虑使用 CRF 条件随机场优化边缘轮廓；针对细胞核预测不全的问题，考虑使用降低候选区域筛选标准提高模型召回率；针对模型无法处理“欺诈”图像（无细胞核），限制预测细胞核形态，减少错误预测的超大超小型细胞核；针对算法复杂且数据量大这一挑战，考虑将算法训练及预测拓展到大规模集群上，缩短算法的迭代更新时间及加快预测图像的速度。

精准的细胞核分割算法有非常广泛应用前景，将算法集成软件嵌入全片数字扫描仪中，可以在切片扫描成数字图像的同时，预先完成细胞核的分割任务，辅助病理医师更好的进行诊断，解放病理医师，让病理医生更多地关注诊断本身，而不是进行诊断所需的服务性任务。同时也加快了病理诊断速度，一定程度上填补国内病理医师缺口大的问题（全国有执照的病理医生只有 1 万人，缺口高达 9 万人）。

参考文献

- [1] Schüffler P J, Fuchs T J, Ong C S, et al. TMARKER: A free software toolkit for histopathological cell counting and staining estimation[J]. *Journal of pathology informatics*, 2013,4(Suppl).
- [2] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database: *Computer Vision and Pattern Recognition*, 2009. CVPR 2009. IEEE Conference on, 2009[C]. Ieee.
- [3] Lin T, Maire M, Belongie S, et al. Microsoft coco: Common objects in context: *European conference on computer vision*, 2014[C]. Springer.
- [4] Kumar N, Verma R, Sharma S, et al. A dataset and a technique for generalized nuclear segmentation for computational pathology[J]. *IEEE transactions on medical imaging*, 2017,36(7):1550-1560.
- [5] Jia X, Song S, He W, et al. Highly Scalable Deep Learning Training System with Mixed-Precision: Training ImageNet in Four Minutes[J]. *arXiv preprint arXiv:1807.11205*, 2018.
- [6] Otsu N. A threshold selection method from gray-level histograms[J]. *IEEE transactions on systems, man, and cybernetics*, 1979,9(1):62-66.
- [7] Xue J, Titterton D. t-tests, F-tests and otsu's methods for image thresholding[J]. *IEEE transactions on image processing*, 2011,20(8):2392-2396.
- [8] Yang X, Li H, Zhou X. Nuclei segmentation using marker-controlled watershed, tracking using mean-shift, and Kalman filter in time-lapse microscopy[J]. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 2006,53(11):2405-2414.
- [9] Oyeboode K O, Du S, van Wyk B J, et al. Investigating the Relevance of Graph Cut Parameter on Interactive and Automatic Cell Segmentation[J]. *Computational and Mathematical Methods in Medicine*, 2018,2018.
- [10] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015[C].
- [11] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation: *International Conference on Medical image computing and computer-assisted intervention*, 2015[C]. Springer.
- [12] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014[C].
- [13] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks: *Advances in neural information processing systems*, 2015[C].
- [14] He K, Gkioxari G, Dollár P, et al. Mask r-cnn: *Computer Vision (ICCV)*, 2017 IEEE International Conference on, 2017[C]. IEEE.
- [15] MacQueen J. Some methods for classification and analysis of multivariate observations: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1967[C]. Oakland, CA, USA.
- [16] Coelho L P, Shariff A, Murphy R F. Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms: *Biomedical Imaging: From Nano to Macro*, 2009. ISBI'09. IEEE International Symposium on, 2009[C]. IEEE.
- [17] Naylor P, Laé M, Reyat F, et al. Nuclei segmentation in histopathology images using deep neural networks: *Biomedical Imaging (ISBI 2017)*, 2017 IEEE 14th International Symposium on, 2017[C]. IEEE.
- [18] Jaccard P. The Distribution of the Flora in the Alpine Zone[J]. *New Phytologist*, 2010,11(2):37-50.
- [19] Sirinukunwattana K, Pluim J P W, Chen H, et al. Gland Segmentation in Colon Histology Images: The GlaS Challenge Contest[J]. *Medical Image Analysis*, 2016,35:489.
- [20] Krasin I, Duerig T, Alldrin N, et al. Openimages: A public dataset for large-scale multi-label and multi-class image classification[J]. *Dataset available from <https://github.com/openimages>*, 2016,2(6):7.
- [21] Vahadane A, Peng T, Sethi A, et al. Structure-Preserving Color Normalization and Sparse Stain Separation for Histological Images[J]. *IEEE Transactions on Medical Imaging*, 2016,35(8):1962-1971.
- [22] Lin T Y, Dollár P, Girshick R, et al. Feature Pyramid Networks for Object Detection[J]. 2016:936-944.

-
- [23] Kingma D P, Ba J. Adam: A Method for Stochastic Optimization[J]. Computer Science, 2014.
 - [24] Folk M, Cheng A, Yates K. HDF5: A file format and I/O library for high performance computing applications: Proceedings of supercomputing, 1999[C].
 - [25] Li M, Andersen D G, Park J W, et al. Scaling Distributed Machine Learning with the Parameter Server.: OSDI, 2014[C].
 - [26] Cui Y, Zhang G, Liu Z, et al. A Deep Learning Algorithm for One-step Contour Aware Nuclei Segmentation of Histopathological Images[J]. 2018.
 - [27] Schmidt U, Weigert M, Broaddus C, et al. Cell Detection with Star-convex Polygons[J]. arXiv preprint arXiv:1806.03535, 2018.
 - [28] Kazeem K O. Improved Thresholding Method for Cell Image Segmentation Based on Global Homogeneity Information[J]. Journal of Telecommunication, Electronic and Computer Engineering (JTEC), 2018,10(1):13-16.