

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(2023)04-1119-15

论文引用格式: Wen K, Jin X, An H, He J and Wang J. 2023. CentroidNet: a light-weight, fast nuclei centroid detection model for breast Ki67 scoring. Journal of Image and Graphics, 28(04):1119-1133(文可, 金旭, 安虹, 何杰, 王珏. 2023. CentroidNet:轻量快速的乳腺癌 Ki67 细胞核中心点检测模型. 中国图象图形学报, 28(04):1119-1133)[DOI:10.11834/jig.211207]

CentroidNet: 轻量快速的乳腺癌 Ki67 细胞核中心点检测模型

文可¹, 金旭¹, 安虹^{1*}, 何杰², 王珏²

1. 中国科学技术大学计算机科学与技术学院, 合肥 230026; 2. 中国科学技术大学附属第一医院病理科, 合肥 230031

摘要: 目的 Ki67 分数是乳腺癌预后评估的重要指标, 计算该分数的关键步骤是检测阴性与阳性癌细胞核。人工检测面临疲劳与主观差异的问题。卷积神经网络有望实现高质量、自动化的细胞核检测, 然而需要病理专家为其标注细胞核。为了减轻标注的工作量, 不少研究者提出以中心点标注训练卷积神经网络。然而这些方法采用过于复杂的卷积神经网络和后处理流程, 未能充分提高易用性和效率、发挥卷积神经网络的质量。对此, 提出 CentroidNet 模型, 旨在提高中心点检测的质量、效率和易用性。**方法** CentroidNet 模型在图像上放置均匀排布的锚点, 为每个锚点预测一个候选点, 一部分候选点通过基于阈值的筛选策略成为预测点。本文提出最近锚点匹配策略用于生成训练标签, 既保证了端到端推理, 又规避了其他一对一标签匹配算法所具有的标签抖动问题。本文建议锚点间距应尽可能接近训练集答案点间最短距离的第一百分位数, 并指出这样的锚点间距能够在前景标签数、坐标回归难度与效率之间取得良好的平衡。本文在设计卷积神经网络的结构时, 没有采纳广为使用的 U-Net 或特征金字塔(feature pyramid network, FPN)中的多级上采样与旁路连接, 反而提高了质量和效率。**结果** 本文在 BCDData 数据集上评估 CentroidNet 模型的质量与效率。BCDData 是目前规模最大的、公开的乳腺癌 Ki67 癌细胞核中心点检测数据集。在质量方面, CentroidNet 取得的综合 F1 分数为 0.879 1, 媲美当前的最高质量。在效率方面, CentroidNet 的推理速度为 12.96 ms/幅、显存占用为 138.8 MB/幅, 达到了当前最高的效率, 远低于若干主流或最新的模型。**结论** CentroidNet 具有高质量、高效率和高易用性; 与现有同类模型相比, 进一步提高了乳腺癌 Ki67 细胞核中心点检测的可行性。

关键词: 乳腺癌; Ki67 分数; 中心点检测; 一对一标签匹配; 锚点

CentroidNet: a light-weight, fast nuclei centroid detection model for breast Ki67 scoring

Wen Ke¹, Jin Xu¹, An Hong^{1*}, He Jie², Wang Jue²

1. School of Computer Science and Technology, University of Science and Technology of China, Hefei 230026, China;

2. Department of Pathology, The First Affiliated Hospital of University of Science and Technology of China, Hefei 230031, China

Abstract: **Objective** Breast cancer-prognostic Ki67 score can be as a key indicator for the proliferation rate of malignant (invasive) cells. Negative and positive nuclei detection is an essential part of Ki67 scoring. An automated algorithm for nuclei detection can alleviate the negative impact of intra/inter-observer variation and labor-intensive nuclei counting. In

收稿日期: 2022-01-04; 修回日期: 2022-05-06; 预印本日期: 2022-05-13

* 通信作者: 安虹 han@ustc.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助(YD2150002001)

Supported by: Fundamental Research Funds for the Central Universities(YD2150002001)

recent years, deep learning methods have been developing intensively in relevant to recognition tasks on pathology images in terms of their learning potentials. Deep learning-based model can learn features derived from the raw data and relieve labor-intensive annotation of training images for pathologists. To alleviate the labor cost of annotation, our research is focused on modeling Ki67 nuclei detection as the centroid detection problem. The existing centroid detection models are commonly used to convert the centroid annotation into a probability map with the same size as the input image, the centroid detection problem is thus converted to a semantic segmentation-like problem. However, a semantic segmentation model has restricted by huge computation cost due to its huge amount of convolutional layers-related decoder. Nevertheless, due to the non-deep-learning post processing on the output heatmap, the whole detection process is complicated and inefficient, and the quality of a deep learning model is required to be developed further. Our CentroidNet model is facilitated to optimize nucleus centroid detection. **Method** The CentroidNet consists of a fully convolutional centroid detector initiated by ResNeXt, along with the detector's training and inference methods. The CentroidNet can place evenly spaced anchor points on the input image. Actually, the anchor points are the places where the center of the detector's receptive field is sited on. Each anchor point-interconnected detector can predict the classification probabilities and the offset to the anchor point, which is called a candidate point as a whole. In this way, the anchor points are the niches of the candidate points. A candidate point has the highest probability and is higher than 0.9. It is treated as predicted point. To assign labels for anchor/candidate points, we implement the "nearest anchor" strategy. This strategy can assign any annotated point to its nearest anchor point literally. Those anchor points are not assigned via any background label-assigned annotated point, whereas those classified label is "background" and label-coordinated is the anchor's coordinate. This strategy is mutual-benefited between annotated points and anchor/candidate points. The label jittering problem can be avoided, which is often seen in current one-to-one assignment strategies. The anchors-between spacing is concerned about more. We recommend that the anchor spacing approximate the First Percentile of the minimal distance between each annotated point in the training set to its neighboring annotated points. Such spacing can balance the ratio of foreground labels, regression-coordinated, and optimal efficiency. The CentroidNet's detector can avoid the commonly adopted paradigm of U-Net or feature pyramid network (FPN), which involves shortcut connections and multiple upsampling layers. This lightweight detector can improve quality and efficiency. **Result** We evaluate the quality and efficiency of CentroidNet model on BCData, the largest publicly available dataset for centroid detection of Ki67 carcinoma nuclei in breast cancer. For quality analysis, the CentroidNet can achieve an averaged F1 score of 0.879 1 compared to the SOTA (state of the art) score. For efficiency evaluation, the CentroidNet can achieve SOTA with an inference speed of 12.96 ms/image and a GPU memory footprint of 138.8 MB/image. **Conclusion** The CentroidNet is featured of light-weight, efficient and ease of use. It has the potentials for centroid detection of Ki67 carcinoma nuclei in breast cancer.

Key words: breast cancer; Ki67 score; centroid detection; one-to-one label assignment; anchor point

0 引言

对乳腺癌患者来说,科学的预后分析能够减少不必要的药物暴露(往往有较大的毒副作用)和医疗费用支出,进而提高患者的生活质量。在乳腺癌的预后分析中,Ki67分数反映了恶性细胞的增殖率,是一项不可忽视的指标(Dowsett等,2011;Yerushalmi等,2010)。Ki67评分的重要环节是识别阴性与阳性的癌细胞核。该环节受到3方面现实因素的制约。1)病理医生需要观察大量的、密集的细胞核,难免产生疲劳和误差;2)我国病理领域严重缺乏人力

资源(卞修武和平轶芳,2019);3)病理医生之间、一位病理医生的不同时期不可避免地具有主观差异(Elmore等,2015)。利用计算机视觉算法来分析Ki67图像,能够使细胞核识别环节变得自动化,将病理医生从繁重的切片阅读工作中解放出来。近年来,卷积神经网络(convolutional neural networks, CNN)极大推动了计算机视觉的发展(LeCun等,2015)。这一趋势激励了不少研究者用CNN解决病理相关的视觉识别问题,并普遍取得了超越手工设计(hand-crafted)方法的识别质量(Srinidhi等,2021)。CNN具有强大的拟合与表征能力,使得其能够直接地从原始图像提取特征并做出决策,从而

减少或消除手工设计的方法。

然而, CNN 在分担病理医生的工作之前, 又带来新的工作, 即标注训练图像。这无疑会降低病理医生参与开发 CNN 的积极性。为了将 CNN 用于乳腺癌 Ki67 细胞核检测, 本文首先需要考虑采用什么样的标注格式。在标注每一个阴性与阳性癌细胞核的前提下, 标注的格式可以是细胞核的掩膜(轮廓)、边界框或中心点。显然, 中心点标注所需的工作量是最少的。在乳腺癌 Ki67 细胞核计数这个特定的任务中, 细胞核的尺寸与形状只是病理医生判别细胞核类型的隐变量, 并不需要记录在案。考虑到 CNN 具有强大的学习能力, 本文认为, 即使只标注中心点、不标注隐变量, CNN 也能够学习到细胞核的其他特征。

事实上, 不少研究者已经打通了利用细胞核中心点标注来训练 CNN 的技术路线。例如, Qu 等人(2020)利用 K-means 和 Voronoi 算法将中心点标注加工成细胞核热力图标签与边缘热力图标签, 训练的热力图模型用于 MoNuSeg 数据集(Kumar 等, 2017)上的细胞核分割任务, 达到了媲美全监督方法的分割质量。Tian 等人(2020)在 Qu 等人(2020)方法的基础上引入 Sobel 热力图标签和由粗到精的自激励学习机制, 进一步缩小了与全监督分割模型的质量差距。Lin 等人(2022)在 Qu 等人(2020)方法的基础上引入自监督预训练机制, 代理任务为从苏木精通道重建苏木精—伊红图像。

但以上方法在后处理与 CNN 结构方面尚有不足之处。后处理是这些方法必备的步骤, 目的是从预测热力图提取中心点坐标。而后处理引入较多的超参数, 降低了模型的易用性。CNN 结构通常采用 U-Net 或特征金字塔(feature pyramid network, FPN)风格的旁路连接与多级 2× 上采样, 本文发现这样的 CNN 结构反而降低了模型在乳腺癌 Ki67 图像上的质量和效率。

为此, 本文提出 CentroidNet 中心点检测模型。该模型具有一个全卷积网络(fully convolutional network, FCN)结构的中心点检测器以及配套的训练与推理流程。检测器以其感受野中心点在输入图像上的停留位置为锚点。当检测器滑动到锚点上时, 预测出最靠近锚点的一个细胞核中心点。

为了在推理阶段简化后处理环节, CentroidNet 在训练时必须采用一对一标签匹配策略(后文简称

一对一策略)。近年来, 一些目标框检测模型采用一对一策略(Wang 等, 2021; Carion 等, 2020; Stewart 等, 2016)。这些模型所采用的一对一策略都是基于线性分配的(后文简称线性分配策略)。线性分配策略大体分为两步。

1) 构造质量矩阵 Q 。该矩阵第 i 行第 j 列元素 $Q_{i,j}$ 是第 i 个答案框与第 j 个候选框的匹配质量, 具体为

$$Q_{i,j} = 1[j \in \Omega_i] \times p_{j,i}^{1-\alpha} \times IoU(\bar{b}_i, b_j) \quad (1)$$

式中, 函数 $1[\text{prop}]$ 返回命题 prop 的真值; 命题 $j \in \Omega_i$ 表示候选框 j 在答案框 i 附近, 详见 Wang 等人(2021)关于空间先验的讨论; $p_{j,i}$ 是候选框 j 在答案框 i 的类别上的概率; $IoU(\bar{b}_i, b_j)$ 是答案框 $\bar{b}_i$ 与候选框 b_j 的交并比(intersection over union, IoU); $\alpha \in [0, 1]$ 是超参数。

2) 求解最优匹配 $\hat{\pi}$ 。具体为

$$\hat{\pi} = \arg \max_{\pi \in \Pi} \sum_i Q_{i,\pi(i)} \quad (2)$$

式中, Π 是一个集合, 枚举了答案框的 n 排列的所有情况, 这里 n 为候选框个数。通常采用匈牙利算法求解式(2)。

线性分配策略使这些模型在推理阶段只需对分类概率进行基于阈值的筛选, 几乎消除了后处理步骤, 从而提高了易用性。然而本文在实践中发现线性分配策略存在标签抖动问题, 即候选点所匹配的答案点索引随候选点本身的变化而变化, 从而干扰损失函数的收敛。考虑到锚点是固定的, 能够与答案点构建稳定的匹配关系, 本文提出最近锚点策略以规避这个问题。

锚点间距对模型质量和效率有着重要的影响。本文建议锚点间距应尽可能接近训练集的常数 $D_{1\%}$, 该常数是训练集的答案点与其他答案点(不区分类别)的最短距离的第一百分位数。本文通过实验揭示出这样的锚点间距能够在前景标签占比、坐标回归难度与效率三者之间取得良好的平衡。

在检测器的结构方面, 本文不使用上采样层与旁路连接, 在质量方面反而超越了大部分使用二者的模型, 并媲美最佳质量; 在效率方面远远超越使用二者的模型。

综上, 本文的贡献如下: 1) 提出最近锚点策略用于训练阶段, 从而简化推理阶段的后处理环节, 进而

提高中心点检测的易用性;并规避现有一对一策略的标签抖动问题。2)建议锚点间距接近训练集常数 $D_{1\%}$,以平衡前景标签占比、坐标回归难度与效率。3)不使用多级上采样与旁路连接,反而提高了中心点检测的质量和效率。

1 本文方法

1.1 CentroidNet 整体的实施流程

1.1.1 推理的整体流程

CentroidNet 推理的整体流程如图 1 所示,实际的锚点间距比图中小。其中,圆点为锚点,圆圈为阴

性预测点,加号为阳性预测点。首先在训练图像上放置锚点阵列 $V^{(0)}$,阵列元素形如 $v^{(0)} = (y^{(0)}, x^{(0)})$ 。然后利用 FCN 检测器自身的滑动性质访问所有锚点,得到候选点阵列 V ,阵列元素形如 $v = (p_0, p_1, \dots, p_K, y, x)$,其中, $p_0, \dots, p_{K-1}$ 为前景概率, p_K 为背景概率。锚点与候选点是一一对应的。最后,对候选点阵列 V 进行筛选,得到预测点集合 $\hat{V}$,集合元素形如 $\hat{v} = (\hat{k}, y, x)$ 。

对每个候选点 v ,是否通过筛选的步骤如下:1)找到候选点的预测概率最大的类别编号 $\hat{k}$,即 $\hat{k} = \arg \max_i p_i$;2)若 $\hat{k} \neq K$,即候选点不是背景,则进入步骤3);3)若 $p_{\hat{k}} \geq 0.9$,则候选点通过筛选。

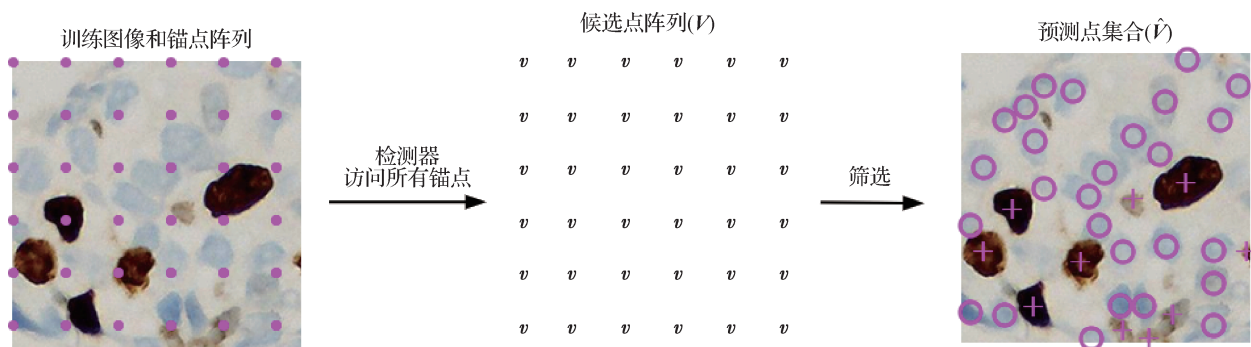


图 1 CentroidNet 推理的整体流程

Fig. 1 The overall process of CentroidNet during inference

1.1.2 训练的整体流程

图 2 以计算图的形式给出了训练的关键流程,即单幅图像的损失函数 $L(V, \bar{T})$ 的计算方式。本文从 3 个视角解读图 2。图 2 中,计算图的节点代表运算,边代表数据。图中实线数据是可导的,虚线数据是不可导的。图中所有大写粗斜体符号都是由相应的小写普通斜体符号组成的阵列,例如 P_0 是 p_0 组成的阵列。

1)获取训练标签 $\bar{T}$ 。点集 A 是病理学家标注的答案点集合,集合元素形如 $a = (k_h, y_h, x_h)$ 。 A 的长度是可变的,不能对齐到检测器的输出,因此需要最近锚点策略将 A 转化为标签阵列 $\bar{T}$,元素为 $\bar{t} = (\bar{k}, \bar{y}, \bar{x})$ 。

2)获取候选点阵列 V 。利用检测器从输入图像 I 得到候选点阵列 V 。注意候选点与标签是一一对应的。

3)定义图像损失函数 $L(V, \bar{T})$ 。图像损失函数

$L(V, \bar{T})$ 是候选点损失函数 $l(v, \bar{t})$ 的和,具体为

$$L(V, \bar{T}) = \sum_{(v, \bar{t})} l(v, \bar{t}) \quad (3)$$

候选点损失函数分为分类与回归两部分,并具有一个自适应的权重。具体为

$$l(v, \bar{t}) = w_k \times (-\log p_k + SL_1(y - \bar{y}) + SL_1(x - \bar{x})) \quad (4)$$

式中,分类部分采用交叉熵(负对数似然)函数,用于衡量检测器的分类误差;回归部分采用 Smooth L1 函数(Girshick, 2015),用于衡量检测器的坐标回归误差。具体为

$$SL_1(x) = \begin{cases} 0.5x^2 & |x| < 1 \\ |x| - 0.5 & |x| \geq 1 \end{cases} \quad (5)$$

同时,考虑到前景标签的数量普遍少于背景标签,本文引入批次内的类别权重机制,目的是增强前景的梯度。类别权重机制反映于式(4)中的权重系数 w_k 。对于 K 类中心点检测任务,本文需要计算 $K+1$ 个这样的系数,具体为

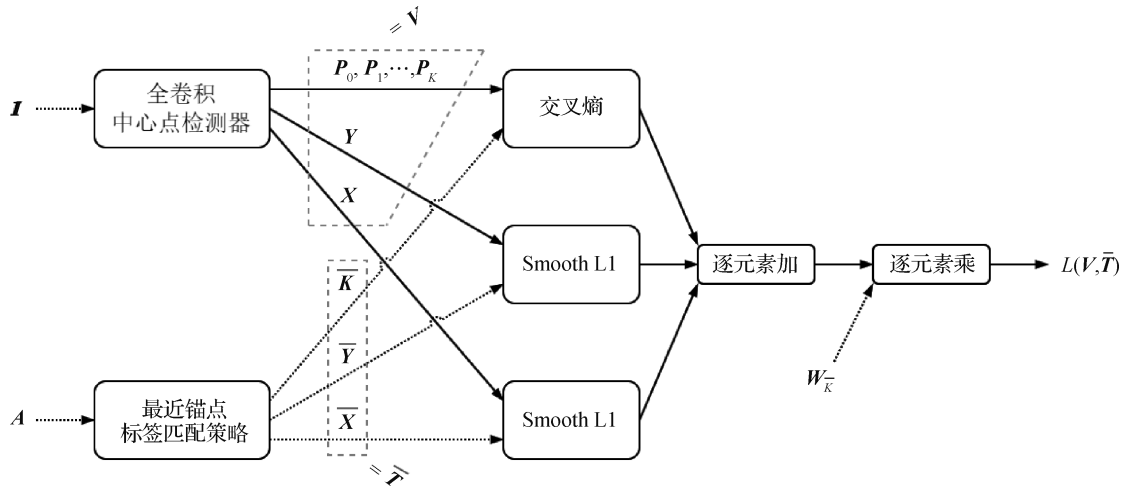


图2 CentroidNet单幅训练图像的损失函数计算图

Fig. 2 The compute diagram of the loss function of one single training image

$$w_k = \sqrt{\frac{1 + \sum_{i=0}^K n_i}{1 + n_k}}, k \in \{0, 1, \dots, K\} \quad (6)$$

式中, n_k 是第 k 类标签的数目, 计数范围是与当前图像同一批次的所有训练标签。

有了图像损失函数, 本文将批次损失函数 L_b 定义为批次内所有图像的 L 的算术平均。之后, 采用批量随机梯度优化器(简称优化器)对 L_b 进行迭代优化, 从而更新检测器的参数。这样的优化器有随机梯度下降法(stochastic gradient descent, SGD)、Loshchilov 和 Hutter(2019)方法等。优化器在每一轮迭代中, 从训练集随机而不重复地抽取一批训练图像 I_b 及对应的答案点 A_b 。当所有训练图像都被访问时, 优化器完成一轮迭代。每一轮迭代之后, 在验证集上评估模型的质量, 并保留该轮训练后的检测器参数 θ_n , n 为该迭代的索引。迭代足够的轮数之后, 选择验证集上质量最高的一轮参数 $\hat{\theta}$ 作为检测器的参数, 在测试集上评估质量与效率。

1.1.3 CentroidNet 检测器

检测器是 CentroidNet 的重要组成部分, 结构如图3所示。其中, ConvKK 代表核大小为 K 的卷积层; MaxPool 为最大池化层; 星号“*”代表步长为2; 逗号后如果跟一个数字则代表输出通道数, 如果跟两个数字则依次代表中间通道数和输出通道数; “(3)”代表连续3个相同配置的模块。除最后一层卷积层外, 检测器中的所有卷积层都依次跟随批量归一化层和 ReLU(rectified linear units)激活层。检测器是 FCN 结构的, 因此检测器在输入图像上从固定的起

点按照固定的步长自动地进行横向和纵向的滑动。这使得检测器感受野的中心点停留在输入图像的固定位置, 本文以这些位置为锚点。

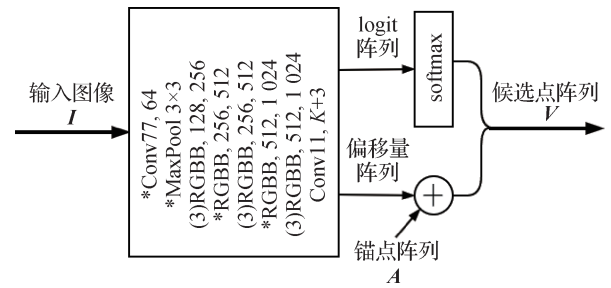


图3 CentroidNet 检测器的结构

Fig. 3 The structure of the detector of CentroidNet

检测器在每一次停留时, 读取输入图像上以锚点为中心点、边长为感受野大小的一块正方形区域, 本文称这一块区域为感受野的内容。之后, 对每一块感受野内容, 检测器计算出对应锚点的 logit 与偏移量, 共 $K+3$ 个元素, 依次为 $u_0, u_1, \dots, u_K, \Delta y, \Delta x$ 。接下来, 利用 softmax 函数将 logit 转化为概率, 并对锚点坐标进行修正, 从而得到候选点 $v = (p_0, p_1, \dots, p_K, y, x)$ 。使用的 softmax 函数和修正锚点坐标的具体计算为

$$p_k = \frac{\exp(u_k)}{\sum_{j=0}^K \exp(u_j)} \quad (7)$$

$$y = y^{(0)} + \Delta y, x = x^{(0)} + \Delta x \quad (8)$$

随着检测器的滑动, 所有锚点都被访问, 从而得到候选点阵列 V 。

RGBB(residual grouped bottleneck block)模块是检测器反复使用的基本组件,结构如图4所示。其中,C1和C2分别为中间通道数和输出通道数,Group是分组卷积的分组数,Res为残差连接。若张量 a 和 b 的通道数相同,则直接进行逐元素相加;否则先利用一个可学习的 1×1 卷积层将 b 的通道数调整到与 a 相同,再进行逐元素相加。RGBB模块的步长为中间 3×3 卷积层的步长。顾名思义,RGBB模块采用了残差连接、分组卷积、瓶颈结构这3个设计要素。其中,残差连接降低深网络的优化难度,避免网络的质量随层数增加而退化(He等,2016);瓶颈结构与分组卷积结合起来,使模型学到更高质量的表征并数倍减少参数数量和运算量(Xie等,2017)。实际上,RGBB模块是ResNeXt网络(Xie等,2017)反复使用的基本组件。考虑到乳腺癌Ki67细胞核类别数远少于ImageNet数据集的1 000类、COCO 2017(common objects in context)数据集的80类,为了在可接受的时间内训练模型,本文所用的RGBB模块的数量仅不到ResNeXt的一半。

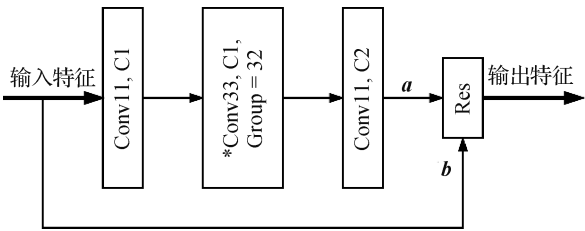


图4 RGBB模块的结构
Fig. 4 The structure of RGBB

1.2 最近锚点策略

本文首先讨论什么样的点适合作为锚点。锚(anchor)是坐标回归的基准位置,是目标检测领域的重要概念。受到FCN目标框检测模型FCOS(fully convolutional one-stage)(Tian等,2019)的启发,本文以FCN感受野的中心点在输入图像上停留的位置为锚点。实际上,这些停留位置为 $(y, x) = (rS, cS)$,其中, r, c 为检测器的输出特征图的行列索引,同时也是锚点阵列的行列索引, S 为检测器的步长。图5形象地展示了锚点在输入图像上的位置。这些锚点组成了锚点阵列 $V^{(0)}$,行数为 $H = \lceil H_0/S \rceil$,列数为 $W = \lceil W_0/S \rceil$, $\lceil \cdot \rceil$ 表示向上取整。

有了锚点的定义,本文能够给出最近锚点策略的定义。该策略首先对每一个答案点,找到与其最

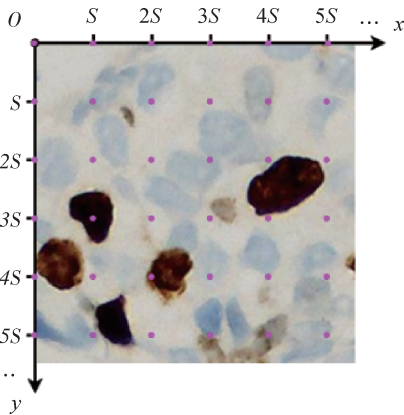


图5 锚点(感受野中心点)出现的位置(紫色圆点)
Fig. 5 Positions where anchor points (or the center point of receptive field) appears (shown as magenta bullets)

近的唯一锚点,令该锚点的标签为答案点;之后对未匹配到答案点的锚点,设置其分类标签为背景、回归标签为锚点自身的坐标。图6展示了最近锚点策略的实施方式。

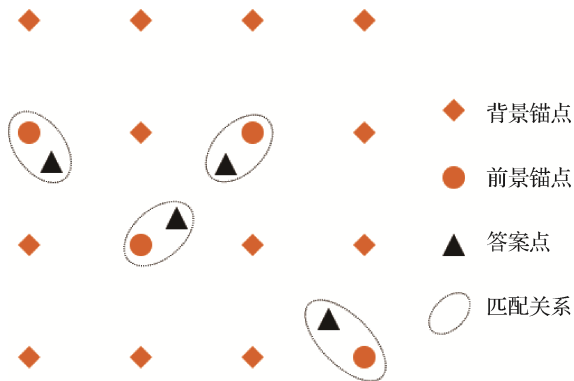
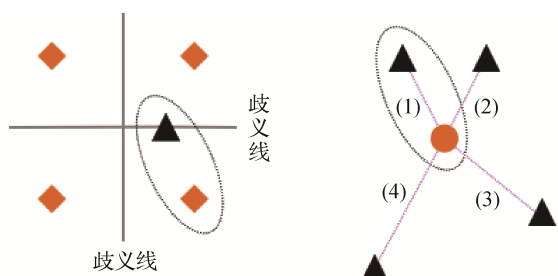


图6 最近锚点策略的实施方式示意图
Fig. 6 The schematic figure of nearest anchor strategy

然而,这样的定义未能覆盖两种偶然的歧义情况。情况1如图7(a)所示,当答案点恰好落在相邻锚点的垂直平分线上时,答案点与这两个相邻锚点的距离是相等的。对于这种情况,本文将答案点匹配到靠下、靠右的锚点。情况2如图7(b)所示,当锚点是多个答案点的最近锚点时,首先取与锚点最近的答案点;如果仍然有多个答案点,再取靠上、靠左的答案点。

按照以上描述,本文仍然不能精确地实施最近锚点策略。为此本文采用更加形式化的语言来定义最近锚点策略。给定一幅训练图像的答案点集 A 和锚点阵列 $V^{(0)}$,最近锚点策略的实施步骤如下:

- 1) 对 A 中的每一个答案点 $a = (k_h, y_h, x_h)$,计算



(a) 一个答案点匹配到多个锚点 (b) 一个锚点匹配到多个答案点

图 7 最近锚点匹配的两种歧义情况及处理方式

Fig. 7 Two ambiguity cases of nearest anchor strategy and its settlement ((a) one annotated point matching many anchor points; (b) one anchor point matching many annotated points)

与之匹配的锚点行号 r 与列号 c , 具体为

$$\begin{aligned} r &= \text{round}\left(\frac{y_h}{S} + 0.5\right) \\ c &= \text{round}\left(\frac{x_h}{S} + 0.5\right) \end{aligned} \quad (9)$$

式中, $\text{round}(x)$ 为对实数 x 进行四舍五入取整。设锚点 $v_{r,c}^{(0)}$ 的标签为 $\bar{t}_{r,c}$, 则 $\bar{t}_{r,c} = a$ 。

2) 对 $V^{(0)}$ 中未匹配到答案点的每个锚点, 设其行号为 i 、列号为 j , 则其标签 $\bar{t}_{i,j}$ 为

$$\bar{t}_{i,j} = (K, y_{ij}^{(0)}, x_{ij}^{(0)}) \quad (10)$$

3) 对 $V^{(0)}$ 中匹配到多个答案点的每个锚点, 设其行号为 i 、列号为 j , 匹配到的答案点序列为 A_q 。对 A_q 进行排序, 排序键(key)依次为答案点到锚点的欧氏距离、答案点的纵坐标、答案点的横坐标, 排序规则为排序键的升序字典序。记 $\hat{a}$ 为排序后的第 1 个答案点, 则 $\bar{t}_{i,j} = \hat{a}$ 。

最近锚点策略保证了任意锚点或候选点的标签是恒定的, 从而规避了预测感知型标签匹配策略带来的标签抖动的问题。而对于最近锚点策略的歧义情况, 首先歧义是小概率事件; 其次本文已经对歧义情况给出了一致的处理方式, 并且可以相信 CNN 的表征能力足以学会这一处理方式。因此本文可以忽略歧义对检测质量的负面影响。

最近锚点匹配策略会将 FCN 检测器训练成滑动的单目标检测器。通过将答案点匹配到最近的锚点, 单目标检测器尝试学习感受野中央的 $S \times S$ 正方形区域内存在哪一类目标的中心点、与中央的偏移量是多少。这为本文接下来调整锚点间距提供了有用的视角。

1.3 调整锚点间距的策略

1.3.1 锚点间距过大或过小带来的影响

锚点间距(即检测器的步长)是 CentroidNet 的一项重要重要的超参数。本文分别列出扩大或缩小步长所带来的事实, 并分析这些事实对质量和效率构成何种影响。

扩大步长带来的事实如下:

事实 1) 前景标签的数目是恒定的, 而锚点总数变少, 从而提高前景标签占比;

事实 2) 对于目标数较多的图像, 锚点数可能少于目标数;

事实 3) 答案点在锚点周围的分布区域变大。

对于事实 1), 在前景标签未达 100% 时, 提高前景标签占比显然是有益的。事实 2) 是事实 1) 的极端情况, 若的确发生, 则必然对质量有害。对于事实 3), 本文将通过迷你实验 1) 来揭示其提高了坐标回归问题的难度, 因而是有害的。

缩小步长带来的事实如下:

事实 4) 答案点在锚点周围的分布区域变小;

事实 5) 锚点总数变多, 提高了背景标签占比;

事实 6) 感受野变小, 使上下文信息变少;

事实 7) 中间特征的宽度与高度增加。

根据迷你实验 1, 事实 4) 能够降低正样本坐标回归的难度, 因而对质量有益。事实 5) 和事实 6) 都有可能对质量有害, 然而本文通过迷你实验 2) 得到一个出乎意料的结论, 当物体大小远低于感受野大小时, 感受野大小对检测质量几乎没有影响, 这与主流的扩大感受野从而提高检测质量的观点是相悖的; 反而是事实 5) 显著地危害检测质量。事实 7) 一定会提高模型的计算量和内存足迹, 是对效率有害的因素。

以上分析只是表明了存在一个步长使检测质量最大化, 并不能精确地给出如何计算这样的步长。在有限的研究条件下, 本文通过实验 3 归纳出 BCDData 数据集 (Huang 等, 2020) 上的最佳步长为训练集的常数 $D_{1\%}$, 并且由于 BCDData 图像的获取条件符合临床场景的要求与习惯, 本文推荐将常数 $D_{1\%}$ 用于大多数乳腺癌 Ki67 图像以及其他类似的图像。

接下来, 本文通过两个迷你实验, 模拟检测器在单个锚点处的“微观”环境与行为, 从而对事实 3)、事实 5) 和事实 6) 给出更有说服力的解释。

1.3.2 迷你实验的数据与方法

在接下来的两个迷你实验中, 本文构造迷你

FCN 检测器,令检测器停留在输入图像的中央处,以模拟 CentroidNet 检测器停留在某个锚点处的环境与行为。检测器识别图像中央的一个正方形区域内的目标类别与中心点坐标。

本文使用方块检测(square detection, SD)数据集来训练迷你 FCN 检测器。SD 数据集由程序生成、白色背景、宽高为 91 像素的 RGB 图像组成。计算机程序(以下简称为生成器)在每幅图像上随机地放置两个宽高为 5 像素的彩色方块,记为方块#1 和方块#2。方块的颜色和分类如表 1 所示。方块具有 13 种颜色。前 12 种颜色被划分为 4 类(编号为 0 到 3)。最后一种是白色,被划为第 5 类(编号为 4),视为背景类。方块#1 被填充为白色的概率为 P_b ,除此之外被填充为任意其他颜色的概率是相等的。方块#2 永远不会被填充为白色,被填充为任意其他颜色的概率是相等的。生成器将图像的训练标签设为方块#1 的类别编号及中心点坐标;当方块#1 为白色时,坐标标签为图像的中央坐标(45, 45)。

表 1 SD 数据集的类别与颜色

Table 1 Classes and colors of SD dataset

类别编号	颜色代码
0	#FF0000, #00FF00, #0000FF
1	#00FFFF, #FF00FF, #FFFF00
2	#800000, #008000, #000080
3	#008080, #800080, #808000
4	#FFFFFF

在使用最近锚点策略的前提下,一个答案点的坐标服从均匀分布,分布空间是一个正方形区域,边长为锚点间距 S ,中心为匹配到的锚点。本文称该区域为锚点的负责区域。方块#1 的目的就是模拟答案点在锚点的负责区域内的分布方式。同时,答案点与其他答案点的最短距离在大部分情况下大于锚点间距 S 且远小于检测器的感受野大小。因此负责区域内出现两个答案点的概率较低,而在感受野内、负责区域外存在其他干扰目标的概率很高。因此本文引入方块#2,用于模拟感受野内、负责区域外的干扰目标。两个方块的位置分布如图 8 所示。方块#1 的中心点随机地出现在以图像中央为中心、边长为 W 像素的正方形区域内部(含边沿)。实际上,方块#1 的中心点坐标服从该正方形区域上的整数

均匀分布。方块#2 的中心点出现在以图像中央为中心、边长为 $W + 2$ 像素的正方形区域外部。实际上,方块#2 的中心点坐标服从其允许区域上的整数均匀分布。此外,若两个方块发生重叠,则重叠部分为方块#1 的对应部分。

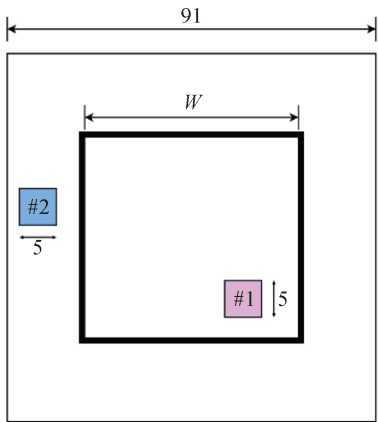


图 8 方块#1 与#2 出现的区域示意图
Fig. 8 Regions where square #1 and #2 present

迷你实验的具体步骤如下:首先,对每一对超参数组合 (W, P_b) ,训练 20 个检测器副本,每个副本训练 2 000 批次;之后,在 200 批次图像(12 800 幅)上,评估每个副本的综合 F1 分数;最后,统计出每一对超参数组合 (W, P_b) 的综合 F1 分数的均值和方差。

由于方块目标过小,本文需对本文正式实验中所述述的 TP 计数规则做出必要的修改。原规则针对乳腺癌 Ki67 图像制订了 10 像素的容许半径;本文针对 SD 数据集将容许半径改为 $\sqrt{2}$ 像素。

1. 3. 3 迷你实验 1:模拟大步长并探究其危害

本文猜想检测器的质量随着锚点负责区域变大而下降。为了验证这一猜想,本文模拟检测器在一个锚点处的环境与行为。为此,本文利用 SD 数据集训练如图 9 所示的单目标检测器,其中 Pool 是最大池化层。该检测器的感受野宽高恰好等于 SD 图像的宽高,感受野的中心恰好是 SD 图像的中心。SD 数据集在本实验中的配置为

$$W \in \{19, 27, 35, 43, 51\}, P_b = 0.2 \tag{11}$$

表 2 给出了负责区域为不同尺寸时,迷你检测器的综合 F1 分数。可以看出,综合 F1 分数随方块#1 分布区域扩大而下降,这说明标签坐标的分布区域越大,坐标回归问题难度越高。本文对此给出进一步的解释。考虑到标签坐标是整数,可以将坐标

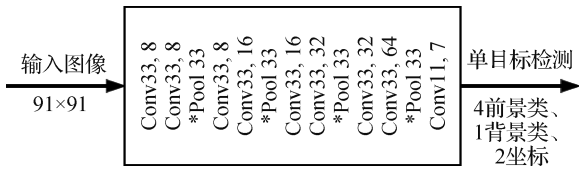


图9 迷你检测器结构图
Fig. 9 The structure of mini detector

表2 迷你实验1的结果

Table 2 The result of mini-experiment 1

类别数	F1	
	均值	标准差
11	0.994	0.001
21	0.966	0.010
31	0.875	0.022
41	0.710	0.034
51	0.413	0.142

注:加粗字体表示最优结果。

回归问题看成 K 分类问题,类别数 $K = (W + 1)^2$,分类标签为整数对 $(\bar{y}, \bar{x})$ 。在这种视角下,分类问题的难度必然与 K 正相关。

综上,迷你实验1的结论为扩大步长对质量的不利影响在于提高坐标的回归难度。

1.3.4 迷你实验2:模拟小步长并探究其危害

本文猜想缩短步长降低质量的主要原因是背景标签变多,而不是感受野变小,并通过本节迷你实验来验证该猜想。SD数据集在本实验中的配置为

$$W = 27$$
$$P_B \in \{0.25/1.25, 1/2, 4/5, 16/17\} = \{0.2, 0.5, 0.8, 0.9412\} \quad (12)$$

式中, P_B 的变化规律是为了模拟CentroidNet的背景标签数量随步长减半而增加到原来的3~4倍的规律。本文通过控制图9中Pool 33层出现与否、出现的位置来改变感受野大小,共有27,47,67,91这4种感受野大小。调整Pool 33层的数量和位置会使检测器输出的特征图宽高不再是1,本文只取特征图中央的向量为检测器的输出。

表3给出了不同感受野大小和背景概率 P_B 下,迷你检测器的综合F1分数。表中,小写字母a,b,c,d依次指代图9从输入端到输出端的4个Pool 33层。aad代表保留a和d两层,并紧跟a层再加一层Pool 33。对于不同的背景概率,越高的F1代表检测问题

表3 迷你实验2的结果

Table 3 The result of mini-experiment 2

感受野大小	P_B	综合F1分数		Pool 33保留情况
		平均值	标准差	
27	0.200	0.985	0.005	cd
47	0.200	0.991	0.005	bcd
67	0.200	0.988	0.003	aad
91	0.200	0.919	0.017	abcd
91	0.500	0.905	0.013	abcd
91	0.800	0.807	0.049	abcd
91	0.941	0.630	0.048	abcd

注:加粗字体表示最优结果。

越简单;对于相同的背景概率,越高的F1代表模型的质量越高。

从表3可以看出,提高背景标签的比例会严重损害模型的质量。当感受野大小为91时,随着背景概率的升高,综合F1分数从0.919降低至0.630。同时,缩小感受野反而提高了质量。当感受野大小从91降至67时,综合F1分数从0.919升至0.988。继续缩小感受野,综合F1分数也能维持在0.988附近。对此,本文的解释是缩小感受野使干扰目标(方块#2)出现的概率降低了。为了验证这一猜想,本文进行了补充实验。利用 $W = 27, P_B = 0.2$ 且没有方块#2的SD数据集,分别测试感受野为27,47,67,91的模型。补充实验中,所有感受野的综合F1分数不小于0.998,方差不大于0.002。补充实验的结果既体现出干扰目标会降低检测质量,也体现出质量与感受野大小的关系较本文预想的复杂。根据既有认知,缩小感受野会减少上下文信息,对模型质量不利。然而缩小感受野也会减少干扰目标的数目,对模型质量有利。本文尚未找到合适的切入点来继续挖掘感受野方面的问题,因此不作更深的探讨。

综上,迷你实验2的结论为缩短步长危害检测质量的主要原因是提高背景标签的比例,而不是缩小感受野。

2 实验结果与分析

为了验证CentroidNet模型的有效性,本文首先验证最近锚点策略的有效性,然后搜索最佳的锚点

间距,最后从质量与效率两方面对比该模型与主流的或最先进的模型。

2.1 实验条件

2.1.1 数据集

实验选用BCData数据集(Huang等,2020)来验证CentroidNet的有效性。BCData数据集是目前为止规模最大的乳腺癌Ki67免疫组化图像数据集,其图像和标注采集自深圳市第二人民医院和中国科学院大学深圳医院的乳腺癌Ki67切片,尺寸为640×640像素。本文在安徽省立医院病理科专家的指导与监督下,利用Adobe Photoshop软件和pytoshop库校准了中心点位置。BCData的划分如表4所示。

表4 BCData的划分情况
Table 4 The divisions of BCData

数据集划分	图像数目	阴性核数目	阳性核数目
训练集	803	69 297	51 124
验证集	133	14 103	7 701
测试集	402	42 450	28 241

为了归纳出可操作性更强的调整锚点间距的方法,本文还测量了BCData训练集每个答案点与其他答案点的最小欧氏距离,记为随机变量 d 。 d 的分布情况如表5所示。

表5 BCData训练集内各答案点与其他答案点的最短距离的分布情况

Table 5 The distribution of the minimum distance from each annotated point to other annotated points in the training set of BCData

百分位	距离 d
0(最小值)	3.5
1	15.3
5	19.0
20	24.0
50(中位数)	30.6
80	39.4
95	51.9
99	70.2
100	228.3

为了避免过拟合,本文对BCData的训练集进行在线随机裁剪、旋转与翻转扩增。裁剪的具体方式

为对图像的每一边裁掉 n 像素, n 服从0~8的整数均匀分布。旋转的具体方式为对图像顺时针旋转 n° , n 服从{0, 90, 180, 270}上的整数均匀分布。翻转的具体方式为随机地进行恒等变换、水平翻转或竖直翻转,各种变换的概率都是1/3。本文的扩增方案将覆盖基线方法原文中的扩增方案。

2.1.2 实验环境

硬件环境:本文的所有训练和推理都是在一台x86_64架构的计算机上开展的。该计算机具有双路Intel Xeon E5-2643 v4中央处理器;具有128 GB主存;具有若干相同型号GPU深度学习加速器,型号为NVIDIA TITAN V,显存为12 GB。

软件环境:本文在Python 3.8上,采用深度学习框架PyTorch(版本号1.10.1)搭建CNN,并结合numpy,scikit-image(skimage),scikit-learn(sklearn)等工具搭建整个训练和推理流程。本文采用Adam(Kingma和Ba,2017)作为所有模型的优化器,初始学习率为1/512,批次大小为8幅图像。CNN参数的初始化策略为xavier_uniform。

一些基线模型在后处理环节需要连通域分析或K-means聚类。为了更公平地展现基线模型的效率,本文采用连通域分析库cc_torch,该库实现了一种前沿的连通域分析算法(Allegretti等,2020);本文采用pykeops.torch库中的K-means算法。以上两个算法库都是基于PyTorch开发的,支持GPU加速。

2.1.3 实验方法

本文采用留出法为每个模型 $F(I; \Theta, H)$ 挑选代表参数 $\hat{\Theta}$ 。其中, Θ 由优化器更新,通常称为参数; H 是人为设定的参数,通常称为超参数。评估一个模型的具体流程分为3步,具体如下:

1) 训练出模型 F 的5个副本,每个副本训练1 500轮。收集这5个副本的最优参数 $\Theta_1, \dots, \Theta_5$,其中 Θ_i 是使第 i 个副本在验证集上取得最高的综合F1分数的参数。

2) 在上述5组最优参数中,记综合F1最高的参数为 $\hat{\Theta}$,即为代表参数。

3) 在测试集上以推理模式运行模型 $F(I; \hat{\Theta}, H)$,记录该模型的质量与效率指标。

2.1.4 质量评价指标

采用查准率(precision, Pr)、查全率(recall, Re)以及F1分数作为模型质量的评价指标。对于有

K 个前景类的中心点检测任务, 本文能够计算出 $3K + 3$ 个质量指标, 分别是任意第 k 类的 $Pr_k, Re_k, F1_k$, 以及综合所有类的 $Pr, Re, F1$ 。

计算这些指标之前, 首先需要对任意第 k 类的 TP_k, GP_k, PP_k 进行计数。这些符号分别代表第 k 类预测正确的预测点数目、答案点数目、预测为第 k 类的预测点数目。本文参考了 Huang 等人(2020)提出的计数算法。若预测点落在以同类答案点为中心、半径 10 像素的圆形区域内, 则该预测点视为一个正确的预测点。然而这个判断标准不够严谨, 最大的问题在于可能有多个同类预测点落在同类答案点的圆形区域内; 其次没有明确地说明单幅图像的指标以何种方式归约到整个数据集的指标。

为此, 本文开发了一种更加严谨的、可以使用 GPU 加速的评价算法。更加严谨的要点在于, 当多个同类预测点靠近同类的某个答案点时, 只有一个预测点计入所在类的 TP 。更快的要点在于, 预测点到答案点的距离是欧氏距离与分类距离的和; 若预测点与答案点同类, 则分类距离为 0; 否则分类距离为图像的对角线长度。

对一幅训练图像, 计数算法的步骤如下:

1) 将预测点集 $\hat{V}$ 转换为独热码矩阵 B 和坐标矩阵 C 。矩阵 B 的第 i 行第 j 列元素等于 1, 当且仅当第 i 个预测点属于第 j 类; 其余元素为 0。矩阵 C 的第 i 行是第 i 个预测点的纵坐标与横坐标。

2) 类似于步骤 1), 将答案点集 A 转化为答案独热码矩阵 $\bar{B}$ 和答案坐标矩阵 $\bar{C}$ 。

3) 对任意前景 k , 令 $TP_k = 0$, 令 GP_k 为矩阵 $\bar{B}$ 的第 k 列元素之和, 令 PP_k 为矩阵 B 的第 k 列元素之和。

4) 求 C 与 $\bar{C}$ 的欧氏距离矩阵 E 。矩阵 E 的第 i 行第 j 列元素是第 i 个预测点与第 j 个答案点的欧氏距离。

5) 求分类错误矩阵 X 。 $X = 1 - B\bar{B}^T$ 。1 为与 $B\bar{B}^T$ 具有相同行数和列数的全 1 矩阵。

6) 求综合距离矩阵 D 。 $D = E + \lambda X$ 。标量 λ 是图像的对角线长度。

7) 从答案点集 A 筛选被命中的答案点集 $\hat{A}$ 。设 $\tilde{V}$ 为访问过的预测点集, $\tilde{V}$ 初始化为空集。对每个答案点实施如下步骤: 设答案点编号为 j , 则找到矩阵 D 的第 j 列的最小元素 $d_{\min}$ 及其行号 $i_{\min}$; 若存在多个最小元素, 取行号最小的元素; 若 $d_{\min} \leq 10$, 则进入

步骤 8); 若集合 $\tilde{V}$ 不存在元素 $i_{\min}$, 则向 $\tilde{V}$ 添加 $i_{\min}$, 同时向 $\hat{A}$ 添加 j 。

8) 遍历 $\hat{A}$ 中每个答案点的类别编号 k , 令 $TP_k \leftarrow TP_k + 1$ 。

有了计数算法, 本文可以定义评价算法。评价算法分为两步。首先通过上述计数算法得到任意第 k 前景类的 TP_k, GP_k, PP_k , 然后计算各类的质量指标。具体为

$$\begin{aligned} Pr_k &= \frac{TP_k}{PP_k} \\ Re_k &= \frac{TP_k}{GP_k} \\ F1_k &= \frac{2 \cdot TP_k}{GP_k + PP_k} \end{aligned} \quad (13)$$

综合所有类的指标为

$$term = \frac{1}{K} \sum_{k=0}^K term_k, \quad term \in \{Pr, Re, F1\} \quad (14)$$

为了方便对比模型的质量, 本文主要以综合 F1 分数为最优先的指标。

2.1.5 效率评价指标

效率评价指标反映在推理时间以及显存足迹两方面。推理时间是模型得到预测点的结束时刻与图像加载到显存的结束时刻的时间差, 单位为 ms/幅。显存足迹是模型得到预测点时的显存占用量与图像加载到显存时的显存占用量之差, 单位为 MB/幅。

2.2 实验 1: 验证最近锚点策略的有效性

最近锚点策略是一种一对一标签匹配策略, 是 CentroidNet 模型在训练阶段的重要环节。为了验证这一环节的有效性, 本文将其与其他一对一标签匹配策略进行对比。目前并没有很多一对一标签匹配策略, 主要有 Wang 等人(2021)和 Stewart 等人(2016)提出的线性分配策略, Wang 等人(2021)提出的贪心策略, 以及本文曾经尝试过的最近候选点策略。在展示实验结果之前, 本文首先说明实现基线方法的要点。

1) 线性分配。为了将该策略迁移到中心点检测任务, 需要对质量矩阵的生成方式做出必要的改动。质量矩阵第 r 行第 c 列元素计算为

$$Q_{r,c} = 1[d_{r,c} \leq d_0] \cdot p_{c,r}^{(1-\alpha)} \cdot \exp(-d_{r,c}^{2\alpha}) \quad (15)$$

式中, $d_{r,c}$ 是第 r 个答案点与第 c 个候选点的欧氏距离; d_0 是距离阈值超参数, 这里取 $d_0 = 10$; $p_{c,r}$ 是第 c 个候选点在第 r 个答案点的类别上的预测概率; α 是

用于调整分类与回归的权重的超参数,这里遵照Wang等人(2021)的做法,取 $\alpha = 0.8$ 。

2) 贪心。贪心使用与线性分配策略相同的质量矩阵,通过求解最优化问题,作为式(2)的快速近似解法。具体为

$$\hat{\pi}(i) = \arg \max_j Q_{ij} \tag{16}$$

3) 最近候选点。最近候选点与最近锚点策略采用相同的匹配规则、消歧义规则,但锚点替换为候选点。

表6展示了4种一对一标签匹配策略在BCData上的质量。可以看到,与3种基线策略相比,最近锚点策略体现出明显更高的质量,综合F1分数分别比线性分配、贪心和最近候选点策略高出0.066 8, 0.076 0, 0.089 8。这是因为3种基线方法的匹配关系随候选点坐标和分类概率的变化而变化。这种不稳定的匹配关系无疑会提高分类与回归问题的难度,从而降低模型的质量。

表6 4种一对一标签匹配策略在BCData上的质量
Table 6 The quality metrics of four one-to-one label assignment strategies over BCData

策略	F1	F1 ₀	F1 ₁
最近锚点	0.879 1	0.869 8	0.888 4
线性分配	0.812 3	0.806 2	0.825 3
贪心	0.803 1	0.793 6	0.812 6
最近候选点	0.789 3	0.780 1	0.798 5

注:加粗字体表示各列最优结果。

2.3 实验2:搜索最佳的锚点间距

锚点间距是CentroidNet的重要超参数。根据迷你实验的结果,锚点间距对质量有着复杂、微妙的影响。然而本文仍然不能精准地找到最优锚点间距。因此本文在BCData上搜索最优锚点间距,并归纳出可操作的结论。

在本实验中,令图3中第0,1,5,9,11模块中的前1~5个的步长为2,其余所有模块的步长为1,从而使锚点间距为2,4,8,16,32。

表7展示了不同锚点间距的质量和效率。不难发现,锚点间距为16时的质量最高,并具有可以接受的效率。这说明锚点间距为16时,模型在前景标签比例、坐标回归难度与效率之间取得平衡。当锚点间距高于16时,坐标回归难度过大,且标签匹配遗漏答案点的概率升高,两个因素都降低质量;当锚

点间距低于16时,前景标签占比过低使质量下降,且隐变量特征图的宽高增加使效率下降。

表7 不同的锚点间距在BCData上的质量和效率
Table 7 The quality and efficiency scores of different anchor spacings (strides) over BCData

锚点间距	F1	推理时间/(ms/幅)
2	0.156 2	237.13
4	0.422 0	59.36
8	0.824 8	19.94
16	0.879 1	12.96
32	0.828 0	10.01

注:加粗字体表示各列最优结果。

考虑到锚点间距16与训练集常数 $D_{1\%} = 15.3$ 极为接近,本文推测 $D_{1\%}$ 是一个合适的锚点间距,并推荐用于未知数据集。假设数据集的细胞核不重叠;那么当 $D_{1\%}$ 较小时,说明细胞核尺寸较小,进而说明目标的信息量较少。此时除了缩小锚点间距至接近 $D_{1\%}$,本文还建议减少检测器层数以适应更少的信息量。类似地,当 $D_{1\%}$ 较大时,本文建议扩大锚点间距至接近 $D_{1\%}$,同时增加层数。

2.4 实验3:与主流或最新模型进行对比

为了验证CentroidNet模型整体的有效性,与其他主流或较新的模型进行对比。参与对比的基线模型包括Lin等人(2022)、LIRNet(local integral regression network)(Zhou等,2021)、Tian等人(2020)、Ribera等人(2019)、FCOS(fully convolutional one stage)(Tian等,2019)和RetinaNet(Lin等,2017)。

实现这些基线模型时,本文模型与对比模型的不同之处在于,Lin等人(2022)模型中自监督学习的代理任务为将H通道重建为H-D(Hematoxylin-DAB)染色图像;Ribera等人(2019)模型中广义均值的参数为 $\alpha = -7$, $\varepsilon = 0.25$,以提高数值稳定性。Lin等人(2022)和Tian等人(2020)模型的细胞核中心点是预测的连通域的中心点。LIRNet和Ribera等人(2019)模型是密度估计模型,为了从预测热力图得到细胞核中心点,本文需对热力图的各通道进行K-means聚类;K-means的样本向量是前景像素的横纵坐标,聚类中心数是模型预测的目标数(四舍五入取整)。FCOS和RetinaNet模型是采用一对多标签匹配的目标框检测模型,FPN只有/4、/8、/16这3个

级别, (m 表示输出特征图的长宽, 是输入图像的 $1/n$) 输出层只有 FPN 的 4 级别; 本文将答案点看成边长为 20 像素、以答案点为中心的正方形框。所有基线模型的概率阈值都是 0.9。为了尽可能公平地对比, 所有基线模型的编码器或骨干网络采用本文图 3 所示的 FCN (不含输出层)。为了配合这样的编码器, U-Net 解码器从 $1/2$ 转为 $1/1$ 的模块与编码器之间没有旁路连接。

表 8 展示了 CentroidNet 与 6 种基线模型的质量和效率。本文对此做出以下 4 点提炼与分析。1) 在质量方面, CentroidNet 的 F1 分数排第 2 名, 但仅比第 1 名模型 (Zhou 等, 2021) 低 0.001 2。本文与第 1 名模型都具有精简的检测器结构。第 1 名模型的检测器虽然采用了 U-Net 风格, 解码器部分却只有两个上采样层。这说明精简的检测器在小规模数据集上

反而具有更高的质量。2) 在效率方面, CentroidNet 的推理时间最短、内存足迹最少, 与第 2 名模型 (Zhou 等, 2021) 相比具有显著优势, 推理时间减少 81.32 ms/幅, 约为第 2 名模型的 $1/7$; 显存足迹减少 47.4 MB/幅, 约为第 2 名模型的 $2/3$ 。能够取得这样的效率, 一是因为 Centroid-Net 检测器与 Zhou 等人 (2021) 模型的检测器相比, 输出宽高更小; 二是 CentroidNet 检测器没有上采样和旁路连接。3) 阳性核的 F1 分数普遍高于阴性核。说明阳性核的识别难度低于阴性核。本文对此的解释是, 阳性核在图像上往往是一整块深棕色的连通域, 内部纹理已无法辨认。4) 在解码器结构方面, FPN 比 U-Net 效率更高, 具有更短的推理时间和更小的显存足迹。这是因为 FPN 的旁路连接为相加, 与 U-Net 的拼接相比, 减少了解码器的通道数。

表 8 CentroidNet 与基线模型在 BCData 上的质量和效率
Table 8 The quality and efficiency scores of CentroidNet and baseline models

模型	F1	F1 ₀	F1 ₁	推理时间/(ms/幅)	显存足迹/(MB/幅)
CentroidNet (本文)	0.879 1	0.869 8	0.888 4	12.96	138.8
Lin 等人 (2022)	0.871 0	0.859 5	0.882 4	135.70	458.6
LIRNet (Zhou 等, 2021)	0.880 3	0.868 2	0.892 3	94.28	186.2
Tian 等人 (2020)	0.863 5	0.851 8	0.875 2	136.43	458.6
Ribera 等人 (2019)	0.814 4	0.811 7	0.817 2	132.91	458.6
FCOS (Tian 等, 2019)	0.868 6	0.857 8	0.879 3	124.46	255.0
RetinaNet (Lin 等, 2017)	0.850 6	0.835 4	0.865 8	124.39	255.0

注: 加粗字体表示各列最优结果。

最后, 本文在图 10 中给出 CentroidNet 的检测样例。其中, 紫色圆圈是答案点, 紫色加号是预测点, 红色箭头指向错误的预测, 红色叉号是错误的答案, 红圈是遗漏的答案。上、下两行分别代表阴性核和阳性核; 左、右两列分别来自测试集图像 #151 和 #161。可以看出, CentroidNet 大体上能够准确而全面地检测乳腺癌 Ki67 图像中的细胞核。然而 CentroidNet 在处理一些难样本时有可能出错。例如, 重叠的两个细胞核可能被识别为一个大细胞核, 褪色 (受限于制片和保存工艺) 的阴性细胞核可能识别不到。此外, 少数错误标注的细胞核也会影响模型的质量。这些错误标注主要集中在弱阳性细胞核。这一类细胞核同时具有棕、蓝两种颜色, 在乳腺癌 Ki67 计数协议中视为阳性, 但受限于人类的视觉疲劳, 在标注

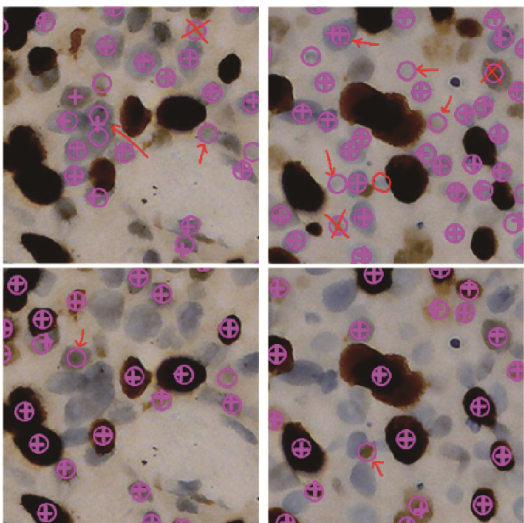


图 10 CentroidNet 中心点检测样例
Fig. 10 Two samples of CentroidNet detection

时有少量被标为阴性。

3 结 论

针对乳腺癌 Ki67 细胞核检测带来的繁重工作及具有较大主观差异的问题,本文提出 CentroidNet 全卷积中心点检测模型。为了减少超参数并提高推理效率,本文将细胞核检测建模为分类—回归问题,在训练模型时采用一对一标签匹配策略。现有一对一标签匹配策略具有标签抖动问题,对此本文提出最近锚点匹配策略,规避了该问题。锚点间距是 CentroidNet 的重要超参数。为了在前景标签占比、坐标回归难度与效率之间取得良好的平衡,本文建议将锚点间距设为训练集答案点间最短距离的第一百分位数 $D_{1\%}$ 。考虑到乳腺癌 Ki67 图像具有较少的目标类别,本文激进地削减了 FCN 解码器(头部)的层数与旁路连接,不仅取得了与当前最高质量相媲美的性能,而且在效率上远超当前最高水平。综上, CentroidNet 在 BCData 上体现出高质量和高效率。受限于难以获取病理图像训练集的现状,本文旨在论证 CentroidNet 用于乳腺癌 Ki67 细胞核检测的可行性,从而提高病理医生在同类图像上制作标注的积极性。

参考文献(References)

- Allegretti S, Bolelli F and Grana C. 2020. Optimized block-based algorithms to label connected components on GPUs. *IEEE Transactions on Parallel and Distributed Systems*, 31(2): 423-438 [DOI: 10.1109/tpds.2019.2934683]
- Bian X W and Ping Y F. 2019. Pathology in China: challenges and opportunities. *Journal of Third Military Medical University*, 41(19): 1815-1817 (卞修武, 平铁芳. 2019. 我国病理学科发展面临的挑战和机遇. *第三军医大学学报*, 41(19): 1815-1817) [DOI: 10.16016/j.1000-5404.201909212]
- Carion N, Massa F, Synnaeve G, Usunier N, Kirillov A and Zagoruyko S. 2020. End-to-end object detection with transformers//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: Springer: 213-229 [DOI: 10.1007/978-3-030-58452-8_13]
- Dowsett M, Nielsen T O, A'Hern R, Bartlett J, Coombes R C, Cuzick J, Ellis M, Henry N L, Hugh J C, Lively T, McShane L, Paik S, Penault-Llorca F, Prudkin L, Regan M, Salter J, Sotiriou C, Smith I E, Viale G, Zujewski J A and Hayes D F. 2011. Assessment of Ki67 in breast cancer: recommendations from the international Ki67 in breast cancer working group. *Journal of the National Cancer Institute*, 103(22): 1656-1664 [DOI: 10.1093/jnci/djr393]
- Elmore J G, Longton G M, Carney P A, Geller B M, Onega T, Tosteson A N A, Nelson H D, Pepe M S, Allison K H, Schnitt S J, O'Malley F P and Weaver D L. 2015. Diagnostic concordance among pathologists interpreting breast biopsy specimens. *The Journal of the American Medical Association (JAMA)*, 313(11): 1122-1132 [DOI: 10.1001/jama.2015.1405]
- Girshick R. 2015. Fast R-CNN//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile: IEEE: 1440-1448 [DOI: 10.1109/iccv.2015.169]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/cvpr.2016.90]
- Huang Z Y, Ding Y, Song G L, Wang L, Geng R Z, He H L, Du S, Liu X, Tian Y H, Liang Y S, Zhou S K and Chen J. 2020. BCData: a large-scale dataset and benchmark for cell detection and counting//*Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention*. Lima, Peru: Springer: 289-298 [DOI: 10.1007/978-3-030-59722-1_28]
- Kingma D P and Ba J. 2017. Adam: A method for stochastic optimization [EB/OL]. [2022-01-04]. <https://arxiv.org/pdf/1412.6980.pdf>
- Kumar N, Verma R, Sharma S, Bhargava S, Vahadane A and Sethi A. 2017. A dataset and a technique for generalized nuclear segmentation for computational pathology. *IEEE Transactions on Medical Imaging*, 36(7): 1550-1560 [DOI: 10.1109/TMI.2017.2677499]
- LeCun Y, Bengio Y and Hinton G. 2015. Deep learning. *Nature*, 521(7553): 436-444 [DOI: 10.1038/nature14539]
- Lin T Y, Goyal P, Girshick R, He K M and Dollár P. 2017. Focal loss for dense object detection//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2999-3007 [DOI: 10.1109/TPAMI.2018.2858826]
- Lin Y, Qu Z Y, Chen H, Gao Z K, Li Y X, Xia L L, Ma K, Zheng Y F and Cheng K T. 2022. Label propagation for annotation-efficient nuclei segmentation from pathology images [EB/OL]. [2022-02-16]. <https://arxiv.org/pdf/2202.08195.pdf>
- Loshchilov I and Hutter F. 2019. Decoupled weight decay regularization//*Proceedings of the 5th International Conference on Learning Representations*. Toulon, France: [s.n.]: <https://openreview.net/forum?id=Bkg6RiCqY7>
- Qu H, Wu P X, Huang Q Y, Yi J R, Yan Z N, Li K, Riedlinger G M, De S, Zhang S T and Metaxas D N. 2020. Weakly supervised deep nuclei segmentation using partial points annotation in histopathology images [EB/OL]. [2022-01-10]. <https://arxiv.org/pdf/2007.05448.pdf>

- Ribera J, Güera D, Chen Y H and Delp E J. 2019. Locating objects without bounding boxes//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 6472-6482 [DOI: 10.1109/cvpr.2019.00664]
- Srinidhi C L, Ciga O and Martel A L. 2021. Deep neural network models for computational histopathology: a survey. *Medical Image Analysis*, 67: #101813 [DOI: 10.1016/j.media.2020.101813]
- Stewart R, Andriluka M and Ng A Y. 2016. End-to-end people detection in crowded scenes//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2325-2333 [DOI: 10.1109/CVPR.2016.255]
- Tian K, Zhang J, Shen H C, Yan K Z, Dong P, Yao J H, Che S, Luo P F and Han X. 2020. Weakly-supervised nucleus segmentation based on point annotations: a coarse-to-fine self-stimulated learning strategy//Proceedings of the 23rd International Conference on Medical Image Computing and Computer-Assisted Intervention. Lima, Peru: Springer: 299-308 [DOI: 10.1007/978-3-030-59722-1_29]
- Tian Z, Shen C H, Chen H and He T. 2019. FCOS: fully convolutional one-stage object detection//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE: 9626-9635 [DOI: 10.1109/ICCV.2019.00972]
- Wang J F, Song L, Li Z M, Sun H B, Sun J and Zheng N N. 2021. End-to-end object detection with fully convolutional network//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 15844-15853 [DOI: 10.1109/cvpr46437.2021.01559]
- Xie S N, Girshick R, Dollár P, Tu Z W and He K M. 2017. Aggregated residual transformations for deep neural networks//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5987-5995 [DOI: 10.1109/cvpr.2017.634]
- Yerushalmi R, Woods R, Ravdin P M, Hayes M M and Gelmon K A. 2010. Ki67 in breast cancer: prognostic and predictive potential. *The Lancet Oncology*, 11 (2): 174-183 [DOI: 10.1016/s1470-2045(09)70262-1]
- Zhou X, Gu M and Cheng Z. 2021. Local integral regression network for cell nuclei detection. *Entropy*, 23 (10): #1336 [DOI: 10.3390/e23101336]

作者简介

文可,男,硕士研究生,主要研究方向为病理图像上的细胞核检测。E-mail: wenke51@mail.ustc.edu.cn

安虹,通信作者,女,教授,主要研究方向为超大规模并行计算芯片和系统结构、高性能计算、医学影像认知计算系统。

E-mail: han@ustc.edu.cn

金旭,男,博士研究生,主要研究方向为智慧病理与医疗大数据。E-mail: jinxu@mail.ustc.edu.cn

何杰,女,主任医师,主要研究方向为临床病理与分子病理学。E-mail: hejie23@ustc.edu.cn

王珏,女,主治医师,主要研究方向为临床病理学。

E-mail: wj_836@163.com