**目 录**

[技术领域 2](#_Toc25085)

[技术背景 2](#_Toc10085)

[完整技术方案 5](#_Toc25751)

[细节处理 7](#_Toc8115)

[模型搭建 7](#_Toc1401)

[特征提取 7](#_Toc329)

[分割模块 8](#_Toc28450)

[分类模块 8](#_Toc1964)

[检测模块 9](#_Toc17915)

[协作模块 9](#_Toc10419)

[模型训练 10](#_Toc18256)

[步骤一：训练分类模块 11](#_Toc30751)

[步骤二：训练分割模块 11](#_Toc23794)

[步骤二：训练检测模块 12](#_Toc12586)

[步骤三：整体协作训练 12](#_Toc13084)

[其他训练细节 14](#_Toc23911)

[数据预处理 14](#_Toc2642)

[数据集划分 14](#_Toc2972)

[模型测试 15](#_Toc3404)

[模型评估 15](#_Toc24197)

[解决的技术问题 16](#_Toc29487)

[有益效果/关键点 17](#_Toc15010)

# 技术领域

涉及医学消化道结肠癌的癌前病变——息肉检测，基于经典预训练模型vgg16进行分类；基于生成对抗方式、带有跳跃连接的类Unet模型进行分割；二者用于指导基于弱监督模型WSDDN进行检测。

# 技术背景

针对自然图像的检测任务，DETR([Carion](https://arxiv.org/search/cs?searchtype=author&query=Carion,+N) et al.2020)、yolo([Redmon](https://arxiv.org/search/cs?searchtype=author&query=Redmon,+J) et al.2016)等方法采用全监督的训练策略使性能有了极大提升，但是将其方法迁移到医学息肉图像上缺十分困难。首先，虽然这些方法能够有效迁移至医学息肉图像的前提是医生在一定量的结肠镜视频中逐帧对息肉进行矩形框的标注。虽然矩形框的标注相比于分割标签（见图1）要简单、容易、快速许多，但依旧需要医生有着丰富的临床诊疗经验，能够完整准确，无重复无遗漏的框出每一帧的所有息肉，并且医生标注的矩形框的准确性将直接影响算法性能好坏的评估。其次，上述模型之所以能有如此好的性能是由于其有标准的矩形框检测标签，标注这些标签耗时耗钱。故弱监督技术应运而生，例如Bilen et al.2016提出弱监督监测模型WSDDN。但对于弱监督检测技术而言，由于不存在标准的标签，模型会不可避免的陷入到局部过拟合的情况。这是由于在检测过程中，模型只能根据图像不同区域的得分不断地从候选区域中筛选出得分最高的区域，这表明得分最高的区域往往是被检测的物体的局部，即物体某些局部特征更具代表性，这一现象在众多检测模型的结果中已经得到验证（见图2）。

图1：矩形框标注 分割标签

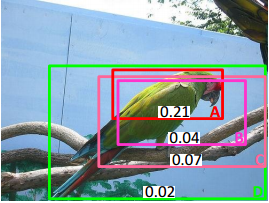


图2检测过拟合问题(Tang et al.2017)

在医学图像息肉的分割任务中，息肉模糊的边界、息肉与背景颜色相近（见图3）造成了息肉分割中的挑战。Ronneberge et al.2015提出了用于医学图像分割的U-Net模型，该模型起初是为了解决细胞层面的分割的任务，但由于模型体积小、速度快、效果好，其被广泛用于解决医学其他图像的分割任务。但由于模型中仅有同层的跳跃连接会使得图像丢失部分全局信息，导致分类精度下降，混淆颜色相近的背景和目标物体。Fan et al.2020提出了平行反向注意模型PraNet用以解决边界模糊问题。实验表明，这些方法确实具有很好的分割效果，但他们都依赖于全监督。这就意味着送入到模型的数要需要医生进行逐像素的细致判断，即对每一个元素进行二分类，判断其属于息肉还是背景。可想而知这种数据标注不仅会耗费大量时间，同时对标注医生的经验有着极高的要求。

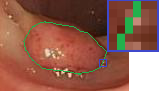


图3息肉边界模糊、息肉与背景颜色相近

息肉的粗略分类有增生性和腺瘤性两种：增生性息肉不易发展成癌症，但腺瘤性息肉却有高达85%的可能恶化成结直肠癌，所以在做结肠镜检查时，腺瘤性息肉必须切除以防止未来的癌症。正是由于如上原因，能否正确区分出这两种息肉对于临床医生和患者而言至关重要。在临床治疗上，医生会根据息肉的种类制定不同的治疗方案，其中腺瘤性患者需要以更短的时间间隔进行密切随访，而增生性患者可以每10年进行一次检查。Patel et al.2020对当下的主流CNN模型进行了息肉分类方面的对比分析，分析表明相较于更新的MnasNet、SE-ResNet、DenseNet等，经典VGG16模型有着更好的性能表现。并且不带归一化操作的VGG16效果反而优于有归一化的VGG16，这可能是由于息肉图像中能用于鉴别息肉类别的像素的精确强度值比自然图像分类中更多，而批归一化层会对像素值进行缩放，影响强度信息，降低性能。

医学息肉图像的检测、分类、分割的单任务确实已经取得了不错的成果，但是他们大多依赖自己任务特定的数据集，不同的任务被独立的研究，这会导致不同任务之间的关联性、互补性以及任务的多样性的缺失。特定的数据集被用于做特定的任务，一旦数据集标注类型发生变化，原任务可能无法进行。此外，在解决不同任务的挑战时，独立的任务只能从自身出发寻找解决办法，往往顾此失彼，没能借鉴其它任务与自己进行取长补短。例如在分割与检测任务中，分割能够很好地覆盖整个实例，但是会将背景与目标混淆；而检测能够很好地区分背景与目标，但很容易过度关注重要区域，如若将二者结合起来则恰好能达到互补的效果，获得更好的任务结果。

我们的模型受Li et al.2019在自然图像上设计的模型SDCN的启发，将其在自然图像中的做法经细节微调迁移应用到医学息肉图像检测领域。本方法在训练时只需要将图像及图像级类别标签作为模型训练的输入，在测试时只需要输入被测试的图片就可给出图像的检测结果。在训练结束后，我们的模型除了能得到优化的检测器外，还能得到优化好的分割器、分类器，在测试阶段对模型进行微调还能同时获得图像的分割与分类结果，但目前我们的模型只关注于检测。

完整技术方案

基于背景技术中所指出的需要大量标注、不同任务之间未考虑合作、检测过拟合、分割混淆背景的技术问题，本发明提供了一种弱监督分割指导检测的2D 息肉检测模型（2D Weakly Supervised Polyp Detection With Segmentation Instruct Detection，WSPD），模型包括四个模块，pC(Polyp Classification)息肉分类模块，pS(Polyp Segmentation)息肉分割模块、pD(Polyp Detection)息肉检测模块、CL(Collaboration Loop)模型协作模块，模型结构如图4所示。

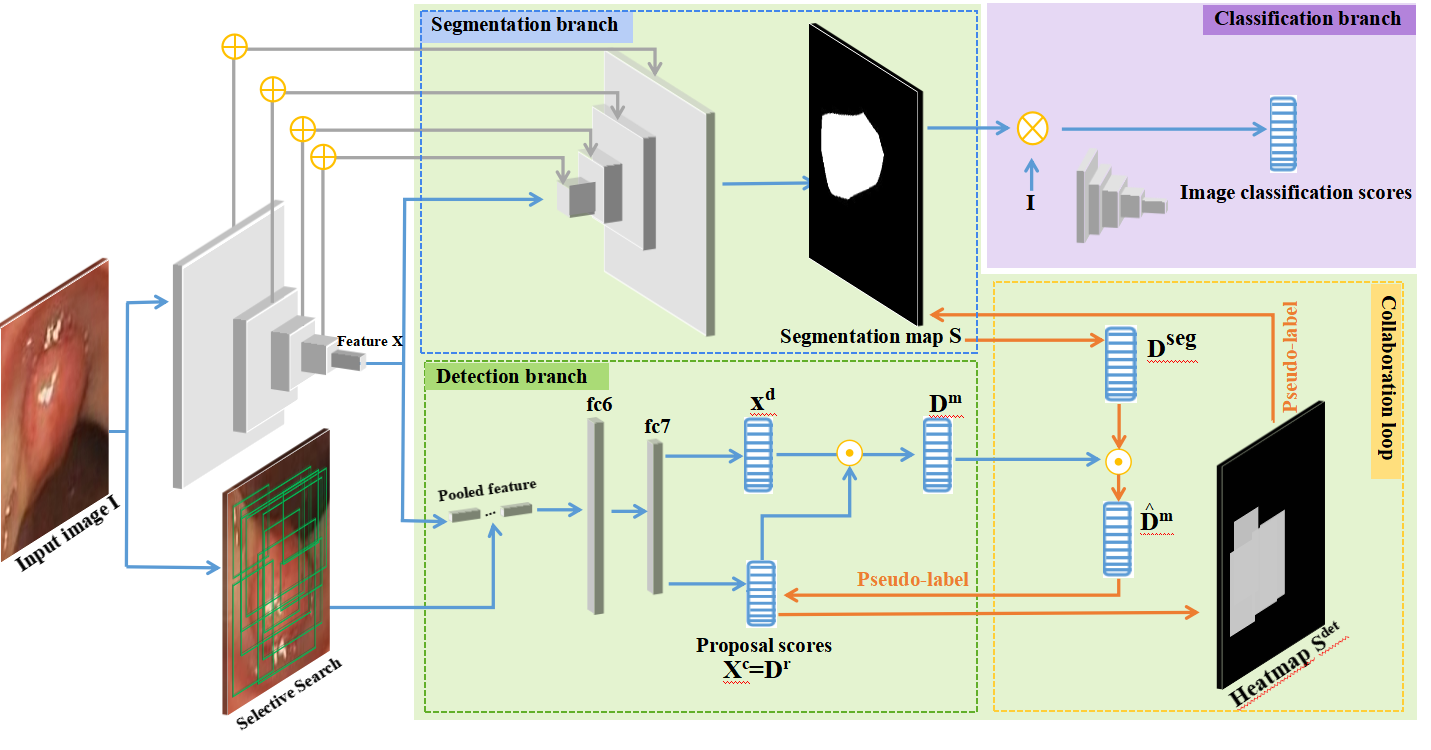


图4 WSPD模型结构图

在训练时对于输入的每张图像I，我们固定其大小为3×224×224（3表示通道数目，224表示图像大小），其特征X由五层卷积网络提取，然后分别送入pS和pD进行分割和检测得到检测结果Dm、Dr与分割结果S，分割结果及其反向结果分别与原图点乘，点乘结果送入pC进行分类，分割与检测结果作为CL模块的输入经CL模块进行协作优化。整个模型由分类标签y = [y1, y2，···，yN]∈{0，1}N(其中N为对象类的数量)监督，整个模型的损失由生成对抗分类损失（）、分类损失（）以及为协作循环提高两个分支准确性而设计的附加协作损失（）。检测指导分割的损失=，分割指导检测的损失=。其中为比例系数分别为：1、0.1、0.1、1、1。在训练中这些损失不断降低，最终达到收敛状态，训练结束。

在测试时，对于输入的每张图像I，I被送入五层卷积网络提取特征，提取到的特征被送入pD模块进行检测最后由Dr给出检测结果，测试结束。

在评价模型性能时，我们采用指标mAP、CorLoc来评价我们的模型，mAP、CorLoc的取值范围为[0，1]，指标值越接近于1模型效果越好。

# 细节处理

## 模型搭建

### 特征提取

我们模型特征的提取以VGG19作为骨架，其中紧挨着第一个maxpool的卷积结果称之为feature1，紧挨着第二个maxpool的卷积结果称之为feature2，紧挨着第三个maxpool的卷积结果称之为feature3，紧挨着第四个maxpool的卷积结果称之为feature4，紧挨着第五个maxpool的卷积结果称之为feature5。具体模型结构如图5中E列所示，其中卷积核大小为3×3，步长为1；采用平均池化，大小为2×2。图像I输入尺寸为[c,h,w]，由于feature1未经池化故其与I相比大小不变，而其他特征都经池化处理所以从feature2开始特征图大小依次变为I的，从feature1开始图像的通道c依次变为64，128，256，512，1024。

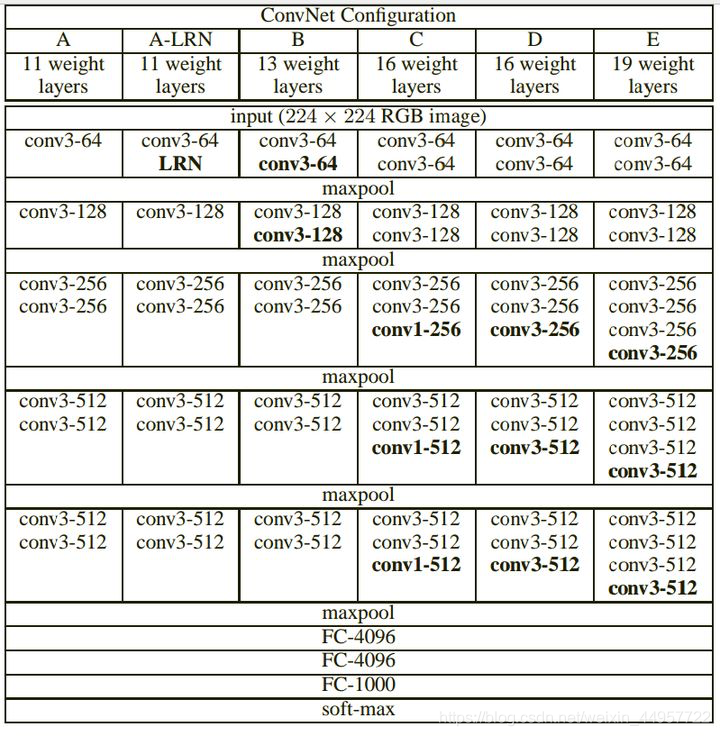


图5 各版本VGG结构图

### 分割模块

分割模块共有5个输入，分别为feature1~4、feature X。分割模块采用同Unet跳跃连接策略：其中feature X经上采样尺寸翻倍、通道数减半与feature4统一尺寸，其结果与feature4进行concatenate得到新的feature4，随后接2层3×3卷积（不改变通道数与大小）整合特征，这使得新feature4与原feature4大小维度一致，但却融合了更深层的图像特征。feature4~1以此类推最后得到尺寸与输入图像I大小一致的特征图。对其再进行最后的卷积操作得到分割图S∈[0，1]（N+1）×h×w，卷积后尺寸大小不变仍与输入图像保持一致，而通道数变为N+1，N为类别数，+1表示背景。通道1~N中每个通道代表一个类别分割结果，N+1通道表示背景类，分割模块结构见图6。

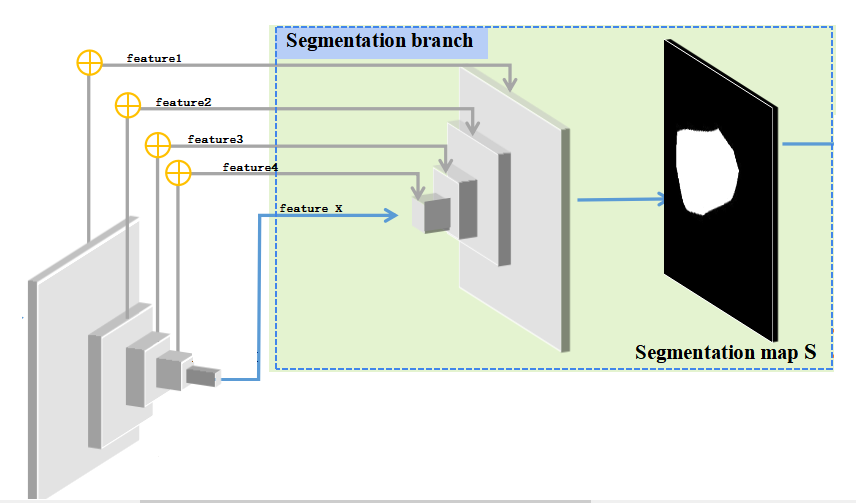


图6分割模块结构图

### 分类模块

我们的模型借鉴已有的研究调查结果，采用分类效果最好的不带归一化的VGG16作为分类模块。VGG16的框架图见特征提取部分图5中C列所示，其中我们只采用其卷积和池化层进行分类的特征提取舍去其所有全连接和softmax，对于分类头而言我们将其设计为N+1分类以此来使用我们的任务。最终分类模块会给出输入数据的类别得分值 ∈ [0，1]N+1 。

### 检测模块

检测模块的输入有2个：1 )提取到的特征feature X、2 )经Selective Search得到B个候选框proposals，其表达形式为左上右下点坐标即(xmin,ymax,xmax,ymin)。由于检测模块的检测头涉及全连接，其需要保证数据拉成向量后大小一致，所以模型首先需要将根据proposals在features5上框出的大小不一致的特征图进行金字塔池化处理，将其统一大小，在我们的模型中我们将其统一为7×7 大小。由于输入的proposals是Selective Search在原图I中提取的，所以在进行金字塔池化操作前还需要对proposals进行映射操作，将其在原图I的坐标映射到特征图feature X。

经上述处理后会得到众多大小一致的特征图，将它们每一个都拉成列向量并在个数的维度上进行堆叠。然后送入两个全连接fc6、fc7，在fc7之后我们的模型将其结果送入两个并行的分支，其中一个分支将各个区域映射到类得分的N+1维向量来执行对各个区域的分类得到得分向量Xc=Dr、令一个分支将各个区域相对于彼此对区域进行评分来执行检测得到结果Xd。即第一个分支预测哪个类与区域相关联，而第二个分支选择哪些区域更有可能包含信息图像片段。最后计算Xc⨀Xd得到每个区域的最终得分Dm，模块结构见图7。

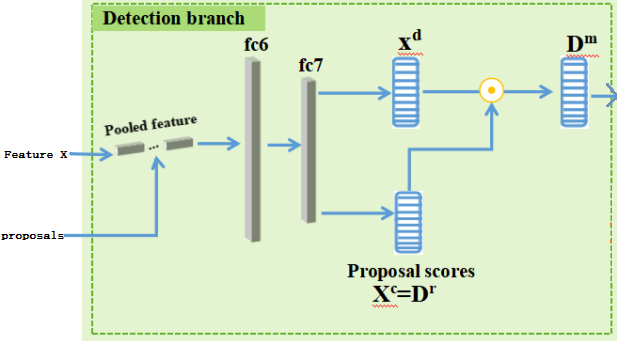


图7 检测模块结构图

### 协作模块

协作模块主要为优化检测分割而设计，分为检测指导分割、分割指导检测。

**检测指导分割：**为了让检测指导分割，首先需要根据检测结果Dr生成检测热图S det∈[0，1]（N+1）×h×w，它可以看作是分割图的模拟。Sdet每个通道c的对应于第c类的热图。具体而言，对于第c个通道，Dr第c列中每个提案框将其分类得分贡献给该提案内的所有像素即：

，

其中表示中点（m，n）、bi表示第i个proposal、Dr（i，c）表示第i个proposal的第c类得分，背景热图可以简单地计算为前景的互补集。为了进一步得到伪类别标签，首先通过最大值的方式离散软分割图S det，保留每个类前60%像素，而忽略其他不明确的像素。

**分割指导检测：**分割指导检测主要通过生成先验概率矩阵Dseg，让其对区域得分xd进行加权即，进而指导分类得分。先验概率矩阵计算公式为：



其中Dseg（i，c）表示第i个对象建议和来自第c通道的分割结果的区域重叠程度、是为增加容错能力而设置的值为0.5的常数。计算完毕后还需将Dseg按列最大值归一化，使其值的范围在[0，1]内。归一化后的Dseg与检测模块区域得分矩阵Dm进行点乘得到加权后的∈[0,1]B×（N+1）,，B为proposals的个数、N+1表示类别。初步得到的经伪标签化操作：对每一类得分最高的proposal及其高度重叠的proposal被标记为图像标签值为1，其余的被标记为背景值为0，得到指导xd的标签Yr（i，c）∈{0,1}B×（N+1）。Yr（i，c）表示第i个proposal的第c个类别。

## 模型训练

我们的模型训练策略同SDCN保持一致，遵循三步训练策略：1）分类模块pC以固定的学习率进行训练，直到其收敛；2） 分割分支pS和检测分支p D在没有协作的情况下被预训练；3） 整个模型结构按照端到端的方式进行训练。

### 步骤一：训练分类模块

我们的模型采用同Patel et al.2020的训练集，与其不同的是我们将类别数设置为三类：增生性息肉、腺瘤性息肉、背景。我们将在ImageNet上训练的VGG16的部分权重作为预训练权重加载至我们的模型中，以0.0001的学习速率训练30次使模型收敛，batch大小设为16，训练损失公式如下：

=

BCE表示标准二进制交叉熵、表示分类模块分类结果、y表示图像I的one-hot标签。

### 步骤二：训练分割模块

图像I经过分割模块得到分割图S，由于不存在像素级伪标签，所以分类模块采用生成对抗思想进行交替训练。分割模块充当生成器、分类模块充当鉴别器。先固定鉴别器以LS训练生成器、再固定生成器以LC训练鉴别器。

**固定鉴别器以LS训练生成器**：分割图S与原图I按S通道进行逐像素乘积运算记为I \* Sc（c∈[1,N+1]），反向分割图adv\_S与原图I按adv\_S通道进行逐像素乘积运算记为I \*（1－Sc）（c∈[1,N+1]）。分割模块的损失目标总体上可以被公式化为每个类别的损失的总和：



为平衡因子分别为1，0.1；为第c个通道的对抗性分类损失、为第c个通道的分类性损失；详细公式描述如下。





其中，∈ {0, 1}N+1，(c) = 1、(i ≠c) = 0；∈ {0, 1}N+1，(c) = 0 (i≠c) = y(i≠c)；pC表示分类模块；avgpool60%sc表示sc的前60%响应像素的平均值；y(c)表示图像第c类的标签值；BCE表示标准二进制交叉熵。其中只有当c为图片正类时才需计算对抗损失。

**固定生成器以LC训练鉴别器**：LC由分类损失和对抗性损失组成。分类模块pC的目标始终是y，因为它不断指导分割挖掘出剩余的对象区域，即使sc被掩盖，具体计算公式如下：



### 步骤二：训练检测模块

图像I经检测模块计算得到每个proposal的最终得分Dm，由于不存在标准的矩形框标签，因此检测分支只能使用仅有的图像级别的标签训练网络。训练由二进制交叉熵损失指导。我们先将区域得分先按列进行累加，结果被视为图像各类别的预测得分，将结果与图像标签计算二进制交叉熵损失，具体公式如下：



其中B表示proposals的个数、N+1表示分类的类别数加背景。

### 步骤三：整体协作训练

模型各模块经上述训练方式进行独立预训练（即完成训练步骤中的1、2步）后，我们的模型将整体按照端到端的方式进行微调训练。整体流程如下（见图8）：

输入图像I经特征提取得到特征图feature X；X被送入pS、pD两个并行的模块进行分割与检测，分别得到分割图S与检测结果Dm、Dr；计算；根据S与Dr生成Sdet和Dseg；计算LD<-S和LS<-D；加和LD<-S和LS<-D作为模型协作损失进行反向传播（此时分类模块固定）；计算分类模块损失LC，固定其余各模块进行反向传播。

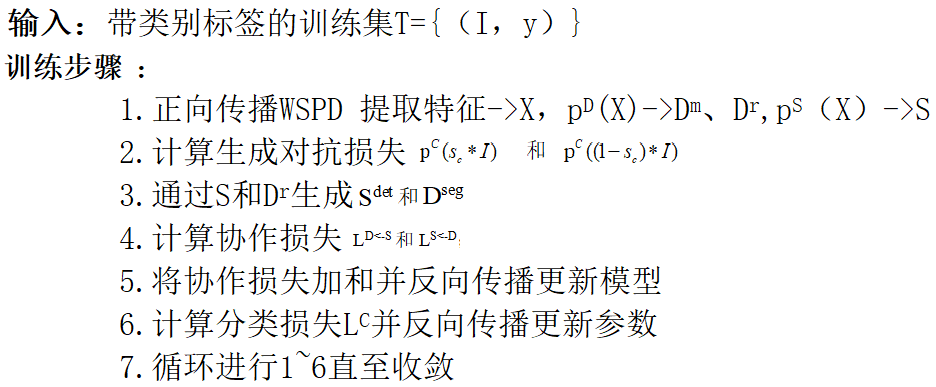


图8 模型协作训练算法

**检测指导分割**时的**分割损失**由两部分构成：单独的分割损失LS以及检测指导下的分割损失。具体公式如下：











为平衡因子0.1，LS与步骤二训练分割模块时保持一致详细内容参见步骤二；是为检测指导分割单独设计的损失，其中LCE表示多分类交叉熵损失，表示伪标签化的Sdet,Sdet的生成参见模型搭建——协作模块，为了给每个像素生成伪类别标签，首先按像素最大值离散Sdet，然后保留每个类的前60%像素,其他像素被视为不明确像素被忽略。

**分割指导检测**时的**检测损失**由两部分损失组成：多目标检测损失以及细化损失，具体公式如下：



为平衡因子均为1；为多目标检测损失其与步骤二训练检测模块时的损失稍有不同：重加权的代替Dm；为了优化检测结果方式过拟合我们特别设计了细化损失，其中LCE表示多分类交叉熵损失，表示对的伪标签化结果，伪标签化策略与模型搭建——协作模块的相同。

**协作时分类损失**与训练分割模块时损失LC保持一致，分类模块pC的目标始终是y，它不断指导分割挖掘出剩余的对象区域，即使sc被掩盖具体具体公式如下：



## 其他训练细节

### 数据预处理

由于数据及图像个数有限，需要对图像进行随机水平/竖直翻转以此来扩充数据集；

### 数据集划分

训练集和测试集的划分比例为训练集80%，测试集20%，由于对数据进行了预处理，同一张图像及其不同扩充图像必须归属于统一数据集（同为训练集或同为测试集），以此确保测试集与验证集不存在数据图像交叉；

## 模型测试

在测试时只需将训练好的特征提取、检测模块的权重参数文件加载到模型中即可。将待测试的图像一张张输入模型得到检测结果，对结果进行NMS后处理得到最终检测结果。

## 模型评估

[在评估模型时我们采用常用的检测模型评价指标mAP(IoU@0.5)、CorLoc。](mailto:在评估模型时我们采用常用的检测模型评价指标mAP(IoU@0.5)、CorLoc。)







其中CorLoc表示正确定位的程度等价于准确率，AP在计算时任选11点法或者积分法；TP：真阳性，与目标框IoU>0.5的检测框数量（同一个Ground Truth 只计算一次）、FP(False Positive)：假阳性，与目标框 IoU<=0.5的检测框、FN（False Negative）：假阴性，没有被检测到的真实目标框（ground\_truth）的数量。

# 解决的技术问题

1. 对于**高标注成本**做视觉任务的问题，我们的方法只需要图像类别，相比于标准的矩形框和掩膜标注而言这种被称为弱标注。在仅有弱标注的情况下，我们的模型能同时完成分割、分类、检测三大视觉任务，这不仅显著降低数据标注成本的，还同时提升了模型的任务效率。对于检测/分割任务需要精细的标注问题，我们的方法只需要图像类别，相比于标准的矩形框和掩膜标注而言这种被称为弱标注。在仅有弱标注的情况下，我们的模型能同时完成分割、分类、检测三大视觉任务，这会显著降低数据标注成本。
2. 对于模型**聚焦于单任务**研究忽略不同任务互补性质的问题，我们的模型将分割、分类、检测三大视觉任务结合在一起。分割与分类通过生成对抗思想相联系，分类指导分割，分割与分类形成对抗，二者在博弈中不断优化自己；检测与分割通过协作机制相互配合，分割指导检测，检测反过来约束分割，二者在相互取长补短中让自身达到更优。这使得我们的模型能够同时执行不同任务，同时还解决了不同任务由于数据标签改变导致的其他任务无法工作的问题，此外我们的模型经训练后除了能得到优化的检测器，还能得到优化的分类器和分割器。
3. 对于检测任务中易出现的**局部过拟合**问题，我们的模型让分割与检测协作，通过分割指导检测解决过拟合问题。由于分割具有能很好分割能够很好地覆盖整个实例的特性，我们的模型让分割结果与检测的候选区域计算重叠度，根据重叠程度对检测模块的得分结果进行加权，使得分较低但包含目标物体大部分区域的矩形框也能被选为检测最终结果解决了检测任务中过拟合的问题。
4. 对于分割任务中易出现的**混淆背景与目标**的问题，我们的模型让分割与检测协作，通过检测指导分割解决混淆背景与目标的问题。由于检测能够很好地区分背景与目标的特性，我们的模型让检测结果与分割结果计算交叉熵损失，在不断迭代中最小化该损失，使得被误分为目标区域的背景区域不断减少，最后的分割区域会全位于检测框内，达到解决混淆背景与目标的问题。

# 有益效果/关键点

1. **极大地降低对标签数据的需求**：缓解了医生标注的负担以及对标注医生经验的较高要求：对于医学图像来说，有标记图像的获取要比自然图像获取更为困难。一方面，医学图像可能会不可避免的涉及病人隐私问题，若想公开数据需要事先询问病人意愿，征得病人同意。除此以外，与标注自然图像相比，医学图像需要大量具有专业知识的医学人员耗费大量的时间才能够完成精准的注释，这会极大增加获取标记数据的成本，也同时增加了医学人员的负担。
2. **通过协作模块解决了检测过拟合问题、分割混淆背景与目标的问题**：在只有图像级监督的检测中由于检测标准标签的缺失，检测没有标准的参照对象，只能根据分类结果试探性学习，这导致检测容易陷入局部过拟合。这种问题在图像检测中极为常见，纠其原因在于检测的择优过程。对于网络而言学习整个大目标的特征比较困难，故随着网络加深和迭代次数的增多网络更倾向于目标的局部，造成过拟合现象。本模型用分割结果的不同通道代表不同的图像中的不同类别，不同通道和其对应的候选框形成的热图计算重叠程度形成加权矩阵Dseg，加权矩阵对按类别打分的矩阵Dm加权得到 。这将会抬高那些虽框住更多目标内容但得分相对较低的框的得分，让其有被选为最终结果的可能。对于分割而言，将分割结果与检测热图计算LCE会让分割网络不断地将分割结果回收到检测结果框内，剔除那些被误当成目标而分割进来的背景区域，达到细化分割的效果。
3. **迁移图像生成的生成对抗思想，仅用图像及标签训练分割**。分割模块为生成器、检测模块为判别器，二者进行生成对抗，使分割模块能在没有标签的情况下分割出图像中的息肉。分割模块在没有标准标签的情况下会试探性的给图像中每个像素分类，最终形成分割图，分割图及其反向图和原图进行点乘，将点乘结果送入到分类模块，分类模块对他们二者进行分类判断并与标签计算损失。在不断迭代中，分割模块效果越来越好，产生的分割图不断地对目标区域进行擦除操作，以达到欺骗分类模块，使其无法正确分类，而分类模块在迭代中也不断提升自己分类的能力与分割模块形成对抗。不断的迭代使二者的损失逐渐降低最终二者均收敛，分割模块能很好的分割出目标区域，分类模块也能准确的判断出目标的类别。
4. **模型经训练后，能得到多个优化后的检测器**：在各种医学图像的视觉任务中，分割、检测与分类都被独立分开，不同的任务对应自己特有的数据集和模型。训练过程中严格的数据标签使得一个模型陷入只能做单任务的局限中，例如做分类任务的数据集在不添加额外标签的前提下不能用来做检测/分割，其他也是一样。对于数据本身而言，只有拥有了任务要求的标注才能被用于特定任务，局限了数据的使用范围；对于研究人员而言，这些有限的医学图像资源没能被更好的挖掘其价值。而在我们的模型中，只需要给出图像级的分类标签即医生只需要判断图像中的息肉是哪种类型（增生/腺瘤）即可，这相对于手动标注极为快速。且在仅有的标签下我们的模型可以同时优化出检测器、分割器、分类器。分割器能给出图像中息肉分割结果的掩膜图(进一步处理还能将掩膜结果标注在真实图像上)；检测器能给出息肉的检测框及框的得分；检测器能给出图像中息肉的分类结果及其类别概率。这不仅提高了数据的使用效率同时给予了主治医生更多维的参考，分类结果能给予医生治疗方案的建议，减少医生在做结肠镜手术时的负担；灵敏的检测能帮助降低漏诊率；准确的分割结果还能加快手术过程减轻患者的痛苦感。