(19)中华人民共和国国家知识产权局

(12)发明专利

(10)授权公告号 CN 108062574 B
(45)授权公告日 2020.06.16

(21)申请号 201711494009.2

(22)申请日 2017.12.31

(65)同一申请的已公布的文献号
申请公布号 CN 108062574 A

(43)申请公布日 2018.05.22

(73)专利权人 厦门大学
地址 361005 福建省厦门市思明南路422号

(72)发明人 纪荣嵘 沈云航

(74)专利代理机构 厦门南强之路专利事务所（普通合伙）35200
代理人 马应森

(51)Int.Cl.
G06K 9/62(2006.01)
G06N 3/04(2006.01)
G06N 3/08(2006.01)

(56)对比文件
CN 104217225 A,2014.12.17,
CN 107273891 A,2017.10.20,
CN 106227836 A,2016.12.14,
CN 103473787 A,2013.12.25,
CN 107203781 A,2017.09.26,
CN 103456027 A,2013.12.18,
WO 2006075594 X,2008.07.03,
Jim Mutch等.Object Class Recognition and Localization Using Sparse Features with Limited Receptive Fields.《International Journal of Computer Vision》.2008,第80卷(第1期),第45-57页.
许鹏飞.基于空间相关性特征的目标识别方法.《中国优秀硕士学位论文全文数据库 信息科技辑》.2011,(第S1期),第I138-1363页.

审查员 万盼盼

权利要求书2页 说明书12页 附图3页

(54)发明名称

一种基于特定类别空间约束的弱监督目标检测方法

(57)摘要

一种基于特定类别空间约束的弱监督目标检测方法。使用候选区域提取算法提取所有训练图像的候选区域;在训练弱监督目标检测器中,提取每一张训练图像的特定类别的像素梯度图,特定类别的像素梯度图反应像素对特定类别的响应,粗略估计目标物体的形状和位置;计算对应候选区域包含目标物体的置信度;把候选区域的置信度引入候选区域分类得分的聚合过程中,包含候选区域的分类得分和候选区域的空间信息;候选区域的空间约束排除背景噪声区域,获得更准确的模型;在训练过程中使用多中心正则化保证模型的学习过程稳定;在测试弱监督目标检测器中,把图像以及对应的候选区域输入模型,模型输出每个候选区域对于每个类别的预测得分。

CN 108062574 B

1.一种基于特定类别空间约束的弱监督目标检测方法,其特征在于包括以下步骤:

1)在训练弱监督目标检测器前,首先使用候选区域提取算法提取所有训练图像的候选区域;

2)在训练弱监督目标检测器中,提取每一张训练图像的特定类别的像素梯度图,特定类别的像素梯度图反应像素对特定类别的响应,因此模型使用特定类别的像素梯度图粗略估计目标物体的形状和位置;

3)根据目标物体的形状和位置的粗略估计结果和候选区域的空间位置的重叠程度计算对应候选区域包含目标物体的置信度;

4)把候选区域的置信度引入候选区域分类得分的聚合过程中,最后的聚合结果同时包含候选区域的分类得分和候选区域的空间信息;候选区域的空间约束排除大量的背景噪声区域,通过学习获得更准确的模型;

5)在训练过程中使用多中心正则化保证模型的学习过程更稳定;

6)在测试弱监督目标检测器中,把图像以及对应的候选区域输入模型,模型输出每个候选区域对于每个类别的预测得分。

2.如权利要求1所述一种基于特定类别空间约束的弱监督目标检测方法,其特征在于在步骤2)中,所述每一张训练图像的特定类别的像素梯度图估计目标物体的粗略形状和位置为:

$$\nabla D^k = \frac{\delta \boldsymbol{y_k}}{\delta I}$$

$$M_{ij}{}^k = \max_{c\in\{0,1,2\}} |\nabla D_{ij}{}^k|$$

其中,$y\in[0,1]^K$表示对应图像的类别预测结果,z^1 表示第1层的特征,z^1就是输入图像本身I,矩阵M^k就是一张输入图像第k个类别的CPG图,通过反向传播算法,进行计算:

$$\nabla D^k = \frac{\delta y_k}{\delta z^l}\frac{\delta z^l}{\delta z^{l-1}}\cdots\frac{\delta z^2}{\delta I}$$

3.如权利要求1所述一种基于特定类别空间约束的弱监督目标检测方法,其特征在于在步骤3)中,所述根据目标物体的形状和位置的粗略估计结果和候选区域的空间位置的重叠程度计算对应候选区域包含目标物体的置信度,计算每个候选区域的空间密度和上下文区域的空间密度:

$$\rho_{rk} = \frac{1}{\sqrt{|B_r|}}\sum_{i,j\in B_r} 1[M_{ij}^k \geq 0.1\cdot max\, M^k]$$

$$\rho_{rk}^c = \frac{1}{\sqrt{|B_r^c|-|B_r|}}\left\{\sum_{i,j\in B_r^c} 1[M_{ij}^k \geq 0.1\cdot max\, M^k] - \sum_{i,j\in B_r} 1[M_{ij}^k \geq 0.1\cdot max\, M^k]\right\}$$

通过积分图,获得每个候选区域的空间密度和上下文区域的空间密度:

$$ii(i,j) = \sum_{i'\leq i,j'\leq j} 1[M_{i'j'}^k \geq 0.1\cdot max\, M^k]$$

$$\rho_{rk}=\frac{1}{\sqrt{|B_r|}}\{ii(i_2,j_2)-ii(i_1,j_2)-ii(i_2,j_1)+ii(i_1,j_1)\}$$

$$\rho^c_{rk}=\frac{1}{\sqrt{|B^c_r|-|B_r|}}\{\{ii(i^c_2,j^c_2)-ii(i^c_1,j^c_2)-ii(i^c_2,j^c_1)+ii(i^c_1,j^c_1)\}$$

$$-\{ii(i_2,j_2)-ii(i_1,j_2)-ii(i_2,j_1)+ii(i_1,j_1)\}\}$$

其中，ii表示CPG图的积分图，$B_r=\{i_1,j_1,i_2,j_2\}$表示候选区域的坐标，$B^c_r=\{i^c_1,j^c_1,i^c_2,j^c_2\}$表示对应的上下文区域坐标，最后置信度矩阵为：

$$W_{rk}=\rho_{rk}-\rho^c_{rk}$$

$$W_{rk}=\frac{W_{rk}}{\max\limits_{r'<R}W_{r'k}}。$$

4.如权利要求1所述一种基于特定类别空间约束的弱监督目标检测方法，其特征在于在步骤4)中，所述把候选区域的置信度引入候选区域分类得分的聚合过程中：

$$S^+=S\odot W\odot 1[W>0]$$

$$S^-=S\odot W\odot 1[W<0]\odot(-1)$$

$$\boldsymbol{y^+_k}=\sum_{r=1}^{R}S^+_{rk}$$

$$\boldsymbol{y^-_k}=\sum_{r=1}^{R}S^-_{rk}$$

其中，$\odot$表示点对点乘积，S是原来的候选区域得分矩阵，S^+是置信度为正数的候选区域加权后的得分矩阵，S^-是置信度为负数的候选区域加权后的得分矩阵，向量y^+和y^-表示使用累加池化最后得到两个候选区域集合的类别预测结果。

5.如权利要求1所述一种基于特定类别空间约束的弱监督目标检测方法，其特征在于在步骤5)中，所述在训练过程中使用多中心正则化为：

$C=\{c_{km}|0\leqslant k<K,0\leqslant m<M\}$

$$m^*(I,k)=\arg\min_{0\leq m<M}\sum_{z\in Z(I,k)}\|z-c_{km}\|_2^2$$

$$L_{center}=\frac{\sum_I\sum_{k=0}^{K}\sum_{z\in Z(I,k)}\|z-c_{km^*(I,k)}\|_2^2}{2\cdot\sum_I\sum_{k=0}^{K}|Z(I,k)|}$$

其中，c_{km}表示第k个类别的第m个特征中心，$m^*(I,k)$是特征中心选择函数，L_{center}是多中心正则化的损失函数。

一种基于特定类别空间约束的弱监督目标检测方法

技术领域

[0001] 本发明属于计算机视觉技术领域，尤其是涉及一种基于特定类别空间约束的弱监督目标检测方法。

背景技术

[0002] 目标检测是计算机视觉领域中的一个基础性的研究课题，主要需要解决图像里有“什么”和在“哪里”的问题(Papageorgiou,Constantine P.,Michael Oren,and Tomaso Poggio. "A general framework for object detection."Computer vision,1998.sixth international conference on.IEEE,1998.)。近十年来计算机视觉里的目标检测问题得到极大的研究发展，大量基于深度学习的新方法被提出。在现实世界中，不同类别的物体的视觉差异可能是很小的，而同一类别的不同物体的差异不仅受物体物理属性变化的影响，还受成像条件变化的影响。例如，花在生物学上是十分多样的，不同个体间的形状，颜色和纹理等属性是千变万化的。在现实场景中，目标物体往往伴随出现视觉上类似的背景形势，或者目标物体只占据整个场景的很小部分，或者目标物体可能被其它物体遮挡，这些各种可能的情景对目标检测任务构成很大的挑战。我们知道目标检测可以分为两个关键的子任务：目标分类和目标定位。前者回答了图像里有“什么”的问题，后者回答了目标在“哪里”的问题。目标分类任务负责判断图像中是否存在特定类别的目标物体，输出一系列带分数的标签表示特定类别的目标物体出现在图像里的可能性。目标定位任务负责搜索图像中特定类别的目标物体的位置和大小。目标检测有着广泛的实际应用，比如智能视频监控，增强现实、基于内容的图像检索和机器人导航等等。除此之外，目标检测也是很多高级计算机视觉任务的重要前提，比如：身份识别和验证、场景分析和理解等等。综上所述，目标检测无论是在计算机视觉领域里还是在实际应用中，都具有非常重要的意义。因此在最近的二十年里，众多科研人员密切关注目标检测问题并投入大量的精力对其进行研究。而伴随着强大的深度学习和强劲的硬件平台发展，近十年来和目标检测相关的课题和研究不仅有增无减，而且模式多样化，每年都有最新的研究成果发表，最新的实际应用公布。尽管如此，目前目标检测算法的性能 (检测准确率和检测速度)跟人类相比起来还是相差非常远。所以说，目标检测问题没有被完美的解决，依旧是计算机视觉领域里一个重要的、具有挑战性的一个研究课题。

[0003] 通常的目标检测算法是基于有监督学习(Hastie,Trevor,Robert Tibshirani, and Jerome Friedman."Overview of supervised learning."The elements of statistical learning.Springer New York,2009.9-41.)。近年来大多数目标检测相关的研究也是关注基于有监督学习的目标检测算法。除此之外另外一个值得关注的方向是基于弱监督学习的目标检测的研究。弱监督学习(Weakly Supervised Learning)(Torresani, Lorenzo."Weakly supervised learning."Computer Vision.Springer US,2014.883-885.)是机器学习领域和模式识别领域里一个非常热门研究方向。事实上，根据训练数据里监督信息的精细度，可以大致把机器学习划分为三种：有监督学习、弱监督学习以及无监督

学习。根据训练数据里监督信息的形式，弱监督学习其实又可以细分为多示例学习和半监督学习。多示例学习的数据是由若干个只有类别标签的包构成，而每个包包括了若干个没有任何标签的示例。假如一个包里所有的示例至少有一个是正样本，那么这个包的标签是正的。假如一个包里所有的示例都是负样本，那么这个包的标签就是负的。另一方面，半监督学习则是使用少量有监督信息的数据和大量无监督信息的数据一起进行学习的问题。虽然多示例学习和半监督学习是有很大差别的，但是它们都是只需要部分或者不完整的监督信息来进行学习。我们可以看出，弱监督学习是处于有监督学习和无监督学习两个极端的中间。实际上在现实生活中，通常带有弱监督信息的数据是远远多于有监督信息的数据。因此基于弱监督的学习的算法有着广泛和重要的应用场景。但是目前对于基于弱监督学习的目标检测的研究工作还是比较少的。而且基于弱监督学习的目标检测算法的性能也差强人意，因此基于弱监督学习的目标检测是一个十分值得研究的课题。

[0004] 通常训练目标检测需要大量人工标注的精细监督信息：目标类别标签和目标位置标签。目标类别标签通常用只包含0和1的向量来表示，1代表图中存在对应的目标，0表示图中不存在对应的目标。而目标位置标签通常用方形包围盒的形式来表示。通常只需要四个坐标就可以确定一个包围盒。这种精细的目标位置标签通常需要付出大量的人力物力来获取。在人工标注包围盒的过程会还会引入标注偏差进而影响训练结果。事实上，只带有目标类别标签的数据是比较容易获得或者标注的，比如用户在网络上传图像，通常会对图像添加标题或者描述。我们可以从互联网获得大量的弱监督标签信息的数据。因此，一个自然的想法就是只使用只有目标类别标签的数据来训练目标检测器，这也正是本发明要研究的问题。

[0005] 当前基于弱监督学习的目标检测仍存在着严峻的挑战(Oquab,Maxime,等."Is object localization for free?Weakly-supervised learning with convolutional neural networks."Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2015.)。总体来说，基于弱监督学习的目标检测带来的挑战性主要是以下两个 方面：鲁棒性和计算复杂性。

[0006] 类内表观差异和类间表观差异是影响基于弱监督学习的目标检测的鲁棒性的主要原因。目标检测算法的鲁棒性越高，那么目标检测性能的准确率也越高。通常目标检测算法的准确率低的主要原因就是大的类内表观差异和小的类间表观差异导致的。类内表观差异是指同一个类别的不同个体之间的差异，例如，人的不同个体在纹理、姿态、形状、颜色等方面存在差异。实际上，同一个人在不同的图像中看起来也会非常不同，这主要就是由于视角、姿态、背景、光照的变化和遮挡的影响。因此我们可以看出，构建具备泛化能力的表观模型极为困难。这导致基于弱监督目标检测算法相对于基于有监督学习的目标检测算法存在三个明显的问题：第一个问题是算法往往会只会检测到目标最显著部分，从而丢掉其他部分。例如对于动物类别，弱监督目标检测算法往往只检测得到头部，而丢掉身体和四肢部分。这是因为身体部分和四肢有比较大的类内表现差异，而头部却没有那么大的类内表现差异。第二个问题是算法会误认为部分背景区域也是待检测目标的一部分。这是因为很多目标类别往往和特定的背景一起出现。比如船只通常和海水一起出现在图像中，因为算法会误认为水也是船只的一部分。第三个问题是若图像里有多个类别相同的目标，检测器往往会把它们当做是同一个目标。因此大多数基于弱监督学习的目标检测算法无法区分同一

类别的多个目标物体。这主要的因为学习过程中算法没有一个目标或者多个目标的概念。

[0007] 待检测目标类别的数量、类别表观描述子的维度以及待检测目标可能存在的位置和大小是基于弱监督学习的目标检测的计算复杂性主要源自。首先现实世界里有成千上万不同类别的物体。其次类别的表观描述子是高维度的，通常是几千维到上万维。最后待检测目标可能存在的位置和大小的组合也是成千上万的，因此目标检测的计算机复杂性较高，设计高效的基于弱监督学习的目标检测算法至关重要。

发明内容

[0008] 本发明的目的在于提供一种基于特定类别空间约束的弱监督目标检测方法。

[0009] 本发明包括以下步骤：

[0010] 1)在训练弱监督目标检测器前，首先使用候选区域提取算法提取所有训练图像的候选区域；

[0011] 2)在训练弱监督目标检测器中，提取每一张训练图像的特定类别的像素梯度图，特定类别的像素梯度图反应像素对特定类别的响应，因此模型可以使用特定类别的像素梯度图粗略估计目标物体的形状和位置；

[0012] 3)根据目标物体的形状和位置的粗略估计结果和候选区域的空间位置的重叠程度计算对应候选区域包含目标物体的置信度；

[0013] 4)把候选区域的置信度引入候选区域分类得分的聚合过程中，最后的聚合结果同时包含候选区域的分类得分和候选区域的空间信息；候选区域的空间约束排除大量的背景噪声区域，通过学习获得更准确的模型；

[0014] 5)在训练过程中使用多中心正则化保证模型的学习过程更稳定；

[0015] 6)在测试弱监督目标检测器中，把图像以及对应的候选区域输入模型，模型输出每个候选区域对于每个类别的预测得分。

[0016] 在步骤2)中，所述每一张训练图像的特定类别的像素梯度图估计目标物体的粗略形状和位置：

[0017] $$\nabla D^k = \frac{\delta \boldsymbol{y_k}}{\delta I}$$

[0018] $$M_{ij}{}^k = \max_{c\in\{0,1,2\}} |\nabla D_{ij}{}^k|$$

[0019] 其中，$y\in[0,1]^K$表示对应图像的类别预测结果，z^1表示第1层的特征，z^1就是输入图像本身 I，矩阵M^k就是一张输入图像第k个类别的CPG图，通过反向传播算法，进行计算：

[0020] $$\nabla D^k = \frac{\delta y_k}{\delta z^l}\frac{\delta z^l}{\delta z^{l-1}}\cdots\frac{\delta z^2}{\delta I}$$

[0021] 在步骤3)中，所述根据目标物体的形状和位置的粗略估计结果和候选区域的空间位置的重叠程度计算对应候选区域包含目标物体的置信度，计算每个候选区域的空间密度和上下文区域的空间密度：

[0022] $$\rho_{rk} = \frac{1}{\sqrt{|B_r|}}\sum_{i,j\in B_r} 1[M_{ij}^k \geq 0.1\cdot max\, M^k]$$

[0023] $$\rho_{rk}^{c}=\frac{1}{\sqrt{|B_r^c|-|B_r|}}\left\{\sum_{i,j\in B_r^c}1\left[M_{ij}^{k}\geq 0.1\cdot max\,M^{k}\right]-\sum_{i,j\in B_r}1\left[M_{ij}^{k}\geq 0.1\cdot max\,M^{k}\right]\right\}$$

[0024] 通过积分图，获得每个候选区域的空间密度和上下文区域的空间密度：

[0025] $$\mathrm{ii}(i,j)=\sum_{i'\leq i,j'\leq j}1[M_{i'j'}^{k}\geq 0.1\cdot max\,M^{k}]$$

[0026] $$\rho_{rk}=\frac{1}{\sqrt{|B_r|}}\{ii(i_2,j_2)-ii(i_1,j_2)-ii(i_2,j_1)+ii(i_1,j_1)\}$$

[0027] $$\rho_{rk}^{c}=\frac{1}{\sqrt{|B_r^c|-|B_r|}}\{\{ii(i_2^c,j_2^c)-ii(i_1^c,j_2^c)-ii(i_2^c,j_1^c)+ii(i_1^c,j_1^c)\}$$
$$-\{ii(i_2,j_2)-ii(i_1,j_2)-ii(i_2,j_1)+ii(i_1,j_1)\}\}$$

[0028] 其中，ii表示CPG图的积分图，$B_r=\{i_1,j_1,i_2,j_2\}$表示候选区域的坐标，$B_r^c=\{i_1^c,j_1^c,i_2^c,j_2^c\}$表示对应的上下文区域坐标，最后置信度矩阵为：

[0029] $$W_{rk}=\rho_{rk}-\rho_{rk}^{c}$$

[0030] $$W_{rk}=\frac{W_{rk}}{\max\limits_{r'<R}W_{r'k}}。$$

[0031] 在步骤4)中，所述把候选区域的置信度引入候选区域分类得分的聚合过程中：

[0032] $$S^{+}=S\odot W\odot 1[W>0]$$

[0033] $$S^{-}=S\odot W\odot 1[W<0]\odot(-1)$$

[0034] $$\boldsymbol{y_k^{+}}=\sum_{r=1}^{R}S_{rk}^{+}$$

[0035] $$\boldsymbol{y_k^{-}}=\sum_{r=1}^{R}S_{rk}^{-}$$

[0036] 其中，⊙表示点对点乘积，S是原来的候选区域得分矩阵，S^{+}是置信度为正数的候选区域加权后的得分矩阵，S^{-}是置信度为负数的候选区域加权后的得分矩阵，向量y^{+}和y^{-}表示使用累加池化最后得到两个候选区域集合的类别预测结果。

[0037] 在步骤5)中，所述在训练过程中使用多中心正则化为：

[0038] $C=\{c_{km}|0\leqslant k<K,0\leqslant m<M\}$

[0039] $$m^{*}(I,k)=\arg\min_{0\leq m<M}\sum_{z\in Z(I,k)}\|z-c_{km}\|_2^2$$

[0040] $$L_{center}=\frac{\sum_I\sum_{k=0}^{K}\sum_{z\in Z(I,k)}\|z-c_{km^{*}(I,k)}\|_2^2}{2\cdot\sum_I\sum_{k=0}^{K}|Z(I,k)|}$$

[0041] 其中，c_{km}表示第k个类别的第m个特征中心，$m^{*}(I,k)$是特征中心选择函数，L_{center}是多中心正则化的损失函数。

[0042] 本发明是一种新颖的基于特定类别空间约束的弱监督目标检测算法。众所周知，目标检测在计算机视觉领域有着极其重要的地位，也是计算机视觉领域里亟需解决的问题。基于弱监督学习的目标检测和基于有监督学习的目标检测的最主要不同地方在于数据集监督信息的精细程度不同。基于有监督学习的目标检测算法需要带有类别标签和目标物体位置标签的数据集进行训练。而基于弱监督学习的目标检测算法只使用带有类别标签的数据集进行学习。因为类别标签的监督信息量是远远少于位置标签的监督信息量，所以只用类别标签学习的目标检测算法称为基于弱监督学习的目标检测算法。本发明的目标是改进现有的基于弱监督学习的目标检测算法的性能，拉进和基于有监督学习的目标检测算法的差距。在对基于弱监督学习的目标检测算法的研究中，还可以评估出类别标签和目标物体位置标签的监督信息的性价比。从而为目标检测算法寻找出最具性价比的标签，使用性价比高的标签信息，获得性能最佳的目标检测器。

[0043] 本发明提出一种算法探索和结合无监督的目标物体全局的形状和位置信息来协助模型的训练。本发明的主要内容可以概括为以下三点：

[0044] 1.本发明提出特定类别的像素梯度图。在训练过程中，本发明提取图像的特定类别的像素梯度图。基于特定类别的像素梯度图，模型可以粗略估算目标物体的形状和位置；

[0045] 2.本发明利用目标物体的粗略估计和候选区域位置的关系，提出了候选区域的空间约束。基于候选区域的空间约束，模型能把特定类别的全局信息和候选区域的局部信息引入模型的学习过程中；

[0046] 3.本发明提出一种多中心正则化来惩罚预测得分比较高的候选区域的特征和对应类别的特征中心的不一致。多中心正则化使得模型训练更加的稳定。本发明的算法没有提高网络模型的复杂度，也没有使用额外的监督信息。最后，大量的实验结果表明本发明的方法取得了优异的弱监督目标检测和定位性能，并超过目前所有最先进的方法。

附图说明

[0047] 图1为通常的弱监督目标检测方法的框架。

[0048] 图2为WSDDN方法的网络结构。

[0049] 图3为本发明的网络结构。

[0050] 图4为一部分训练图像和对应类别的像素梯度图。

具体实施方式

[0051] 以下实施例将结合附图对本发明作进一步的说明。

[0052] 通常的弱监督目标检测方法的框架如图1所示，通常弱监督目标检测算法的框架和有监督目标检测算法的框架相似：即首先提取出图像中大量的候选区域(region proposal)，然后对这些候选区域进行分类。对于每个类别，候选区域的预测得分越高则表示这个候选区域包含这个类别的目标物体的置信度越高。因此为了使用图像的类别标签作为监督信息训练模型，算法需要把各个区域的分类结果聚合成整张图像的分类结果。最后根据图像分类结果和图像类别标签的误差来学习模型的参数。在弱监督目标检测算法中，常用的聚合方法有最大值池化(max pooling)或者平均值池化(average pooling)。然而这些聚合方法丢失了候选区域的位置信息。也就是聚合过程只考虑每个候选区域的分类得

分，而不考虑它们之间的位置和大小关系。

[0053] 本发明对聚合的过程进行深入的改进，并分别提出特定类别的像素梯度图(Category-Specific Pixel Gradient map)、候选区域空间约束(Region Spatial Constraint)和多中心正则化(Multi-Center Regularization)等方法改进基于弱监督学习的目标检测算法。

[0054] 以下给出具体实施例：

[0055] 首先定义本发明主要使用的符号。这里用$I \in \mathcal{R}^{H\times W\times 3}$表示一张RGB格式的输入图像，$B=\{B_1,B_2...B_R\}$表示对应图像的候选区域集合，$B_i \in \mathcal{R}^4$表示图像上的一个候选区域，$t\in\{0,1\}^K$表示对应图像的类别标签。其中H和W分别表示图像的高度和宽度，R表示对应图像候选区域的数目，K表示数据集的类别数目。同时用$S \in \mathcal{R}^{R\times K}$表示对应图像的目标检测结果，其中第r行第k列表示第r个候选区域正好包含第k个类别物体的预测得分。$y\in[0,1]^K$表示对应图像的类别预测结果。图像类别的预测结果y有正确的类别监督信息t，而候选区域的预测结果矩阵S是没有任何监督信息的。

[0056] 本发明使用WSDDN模型作为模型的基本网络结构(Bilen,Hakan,and Andrea Vedaldi. "Weakly supervised deep detection networks."Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.2016.)。如图2所示，WSDDN是一个双分支的深度卷积神经网络。网路的输入是一张图像I以及此图像的候选区域集合B。候选区域提取的算法有多种，比如selective search(Uijlings,Jasper RR,et al."Selective search for object recognition." International journal of computer vision 104.2(2013):154-171.)和edge boxes(Zitnick,C. Lawrence,and Piotr Dollár."Edge boxes:Locating object proposals from edges."European Conference on Computer Vision.Springer,Cham,2014.)等。图像I经过若干个卷积神经网络的卷积层获得卷积特征图。通常称计算卷积神经网络特征图的若干个卷积层为模型的后端，而在卷积神经网络特征图之后的网络结构称为称模型的前端。在模型的前端固定的时候，我们可以使用不同的后端来获得模型的不同表达能力。不同的模型后端有AlexNet(Krizhevsky,Alex,Ilya Sutskever,and Geoffrey E.Hinton."Imagenet classification with deep convolutional neural networks."Advances in neural information processing systems.2012.)、VGGNet(Simonyan, Karen,and Andrew Zisserman."Very deep convolutional networks for large-scale image recognition."arXiv preprint arXiv:1409.1556(2014).)和GoogLenet(Szegedy,Christian,等. "Going deeper with convolutions."Proceedings of the IEEE conference on computer vision and pattern recognition.2015.)等。通常情况下模型后端的深度越深，模型的表达能力也越强。在获得图像的卷积神经网络特征图后，WSDDN把卷积神经网络特征图和图像的候选区域B一起输入到空间金字塔池化层(SPP)来获得每个候选区域的卷积神经网络特征。每个候选区域的卷积神经网络特征经过两层全连接层FC6和FC7获得每个候选区域的特征向量。每个全连接层后面都有非线性激活ReLU层和DropOut层。在这些层的最后，候选区域的特征向量输入两个不同的分支。

[0057] 第一个分支命名为分类分支。分类分支对每一个候选区域计算每个类别的得分。

候选区域特征向量输入分类分支的全连接层FC8c，全连接层FC8c的输出X^c使用一个SoftMax层进行归一化得到S^c：

[0058] $$S_{ij}^c = \frac{e^{X_{ij}^c}}{\sum_{k=1}^{K} e^{X_{ik}^c}} \quad (1)$$

[0059] 第二个分支称为检测分支。检测分支对每个类别计算每个候选区域的得分。每个候选区域经过全连接层FC8d获得分矩阵X^d。和分类分支类似，得分矩阵X^d也使用SoftMax层进行归一化，得到S^d：

[0060] $$S_{ij}^d = \frac{e^{X_{ij}^d}}{\sum_{r=1}^{R} e^{X_{rj}^d}} \quad (2)$$

[0061] 虽然两个分支有相似的网络结构，但是S^c和S^d的本质区别使得网络能够同时进行分类和检测任务。这两个分支的SoftMax层作用各不一样：其中分类分支的SoftMax层对每个候选区域的不同类别的得分进行归一化，也就是矩阵S^c的每一行元素相加的和为1。检测分支的 SoftMax层对每个类别的不同区域的得分进行归一化，也就是矩阵S^d的每一列元素相加的和为1。X^c、X^d、S^c和S^d的维度是一样：

[0062] $$X^c, X^d, S^c, S^d \in \mathcal{R}^{K \times K} \quad (3)$$

[0063] 最后每个候选区域的得分S是两个分支输出矩阵S^c和S^d的乘积：

[0064] $$S = S^c \odot S^d \quad (4)$$

[0065] 其中$\circ$表示点对点的乘法，即Hadamard乘积。事实上，得分矩阵S就是目标检测所得到的结果。也就是每个候选区域的预测得分是分类分支得分和检测分支得分的结合。本发明根据得分矩阵S对每个类别的所有候选区域进行排序。对于一个类别，若候选区域的得分越高，则这个候选区域越有可能包含这个类别的目标物体。最后，通常的目标检测算法会使用非极大值抑制法(NMS)来排除部分重叠比较大的候选区域来得到最终的检测结果。

[0066] 但是目前，WSDDN只得到候选区域的预测得分，而训练数据的监督信息是图像的类别标签。因此WSDDN最后使用了一个累加池化层来获得最后图像类别的预测：

[0067] $$y_k = \sum_{r=1}^{R} S_{rk} \quad (5)$$

[0068] 也就是把所有候选区域的第k个类别的得分累加起来获得对图像第k个类别的预测结果。由于前面矩阵S^d已经对每个类别的每个候选区域的得分进行了归一化，因此最后聚合的类别得分的范围在0和1之间，即$y_k \in (0,1)$。

[0069] 最后WSDDN使用一个交叉熵损失函数来进行深度卷积神经网络的训练：

[0070] $$L = \sum_{I}\sum_{k}^{K} t_k \log y_k + (1 - t_k)\log(1 - y_k) \quad (6)$$

[0071] 如图2所示，本发明对WSDDN网络结构进行了改进。首先本发明通过图像类别预测结果y提取出特定类别的像素梯度图M。特定类别的像素梯度图M包含了特定类别目标物体形状和位置的粗略估计。特定类别的像素梯度图会在下一节进行详细介绍。结合获得的目标物体粗略的形状和位置估计和候选区B的空间位置信息，可以计算出每一个类别的每个候选区域的置信度矩阵W。我们把置信度矩阵W和得分矩阵S进行相乘，获得最后的每个类别的每个候选区域的得分矩阵。同时根据置信度矩阵W的符号，本发明把候选区域集合分为正例集合和负例集合，以及正例候选区域的得分矩阵S^+和负例候选区域的得分矩阵S^-。最后通过累加池化分别获得y^+和y^-。最后本发明还提出一种新的多中心正则化来使得模型的学习过程更加稳定。

[0072] 本发明方法的流程包括以下步骤：

[0073] 在模型训练前：

[0074] 首先使用候选区域提取算法来提取所有训练图像的候选区域。

[0075] 1)模型训练前，首先使用候选区域提取算法来提取所有训练图像的候选区域。

[0076] 2)在模型训练中，给定一个特定训练图像X_0，可以通过计算模型的一阶泰勒展开，在X_0附近用一个线性函数逼近预测得分y_k：

[0077] $$y_k \approx \omega_k^T X + b_k \quad (7)$$

[0078] 这里ω就是预测得分y_k对输入X在X_0附近的导数：

[0079] $$\omega_k = \frac{\delta \boldsymbol{y_k}}{\delta X}|_{X_0} \quad (8)$$

[0080] 把向量化的输入X换成原来的输入图像I，则第k个类别预测得分y_k对输入图像I的梯度为：

[0081] $$\nabla D^k = \frac{\delta \boldsymbol{y_k}}{\delta I} \quad (9)$$

[0082] 其中$\nabla D^k \in \mathcal{R}^{H\times W\times 3}$，最后通过计算$\nabla D^k$所有通道的最大绝对值获得类别k的像素梯度图：

[0083] $$M_{ij}{}^k = \max_{c\in\{0,1,2\}} |\nabla D_{ij}{}^k| \quad (10)$$

[0084] 在深度卷积神经网络中，梯度∇D^k可以通过使用反向传播算法进行计算：

[0085] $$\nabla D^k = \frac{\delta y_k}{\delta z^l}\frac{\delta z^l}{\delta z^{l-1}}\cdots\frac{\delta z^2}{\delta I} \quad (11)$$

[0086] 其中，z^1表示第1层的特征，z^1就是输入图像本身I。最后矩阵M^k就是一张输入图像第k 个类别的CPG图。

[0087] 3)在模型训练中，过滤掉无用元素的第k个类别CPG图的空间密度为：

[0088] $$\rho_k = \frac{1}{\sqrt{|I|}} \sum_{i,j \in I} 1[M_{ij}^k \geq 0.1 \cdot max\, M^k] \quad (12)$$

[0089] 其中，M_{ij}^k 表示第k个类别的CPG图里第i行第j列的元素。1[ ]表示指示函数，当参数为真时，指示函数返回1，当参数为假的时候，指示函数返回0。CPG图是很稀疏的，大部分元素接近于0。因此本发明使用CPG图的面积的根号作为分母来正则化密度。前面定义了CPG图的空间密度，同样可以用类似方法定义候选区域在CPG图上的空间密度。本发明定义候选区域B_r在 CPG图上的空间密度为：

[0090] $$\rho_{rk} = \frac{1}{\sqrt{|B_r|}} \sum_{i,j \in B_r} 1[M_{ij}^k \geq 0.1 \cdot max\, M^k] \quad (13)$$

[0091] 实际上，可以通过积分图(integral image)的方法来快速有效地计算所有候选区域在 CPG图上的空间密度。首先构建CPG图的积分图：

[0092] $$\mathrm{ii}(i,j) = \sum_{i' \leq i, j' \leq j} 1[M_{i'j'}^k \geq 0.1 \cdot max\, M^k] \quad (14)$$

[0093] 通过以下循环计算来快速获得CPG图的积分图：

[0094] $$\mathrm{ci}(i,j) = \mathrm{ci}(i,j-1) + 1[M_{ij}^k \geq 0.1 \cdot max\, M^k] \quad (15)$$

[0095] $\mathrm{ii}(i,j) = \mathrm{ii}(i-1,j) + \mathrm{ci}(i,j)$ (16)

[0096] 其中，$s(i,j)$是每一行的累加，$\mathrm{ci}(x,-1) = 0$，$\mathrm{ii}(-1,y) = 0$，则计算第k个类别的第r个候选区域$B_r = \{i_1, j_1, i_2, j_2\}$的CPG图的空间密度为：

[0097] $$\rho_{rk} = \frac{1}{\sqrt{|B_r|}} \{ii(i_2,j_2) - ii(i_1,j_2) - ii(i_2,j_1) + ii(i_1,j_1)\} \quad (17)$$

[0098] 直观上说，空间密度ρ_{rk}反应了候选区域在CPG图上的平均像素梯度。若一个候选区域的空间密度ρ_{rk}比较大，则这个候选区域很有可能包含目标物体。若一个候选区域的空间密度ρ_{rk}比较小，则这个候选区域很可能是噪声背景。若直接使用候选区域在CPG图上的空间密度作为对应候选区域的置信度，则会导致只包含目标物体中心的候选区域的置信度太大。因为在CPG图上目标物体中心的空间密度往往会比目标物体边缘的空间密度高很多。因此，加入上下文密度 ρ_{rk}^c 来防止这种情形。通过以下公式计算每个候选区域$B_r = \{i_1, j_1, i_2, j_2\}$的上下文区域$B_r^c = \{i_1^c, j_1^c, i_2^c, j_2^c\}$：

[0099] $$h_c = \frac{i_1 + i_2}{2} \quad (18)$$

[0100] $$w_c = \frac{j_1 + j_2}{2} \quad (19)$$

[0101] $h_r = (i_2 - i_1) \cdot \alpha$ (20)

[0102] $w_r = (j_2 - j_1) \cdot \alpha$ (21)

[0103] $$i_1^c = \max\left(h_c - \frac{h_r}{2}, 0\right) \quad (22)$$

[0104] $$j_1^c = \max\left(w_c - \frac{w_r}{2}, 0\right) \quad (23)$$

[0105] $$i_2^c = \min\left(h_c + \frac{h_r}{2}, \mathrm{H}\right) \quad (24)$$

[0106] $$j_2^c = \min\left(w_c + \frac{w_r}{2}, \mathrm{W}\right) \quad (25)$$

[0107] 这里h_c和w_c分别表示候选区的垂直和水平方向的中心坐标，同时它们也是对应的上下文区域的垂直和水平方向的中心坐标。h_r和w_r分别表示上下文区域的高度和宽度，其中α是缩放因子，本发明设置α＝1.8。最后获得对应的上下文区域的$B_r^c = \{i_1^c, j_1^c, i_2^c, j_2^c\}$。获得上下文候选区的坐标后，计算候选区域的上下文区域在CPG图上的空间密度ρ_{rk}^c：

[0108] $$\rho_{rk}^c = \frac{1}{\sqrt{|B_r^c| - |B_r|}}\left\{\sum_{i,j \in B_r^c} 1\left[M_{ij}^k \geq 0.1 \cdot max\, M^k\right] - \sum_{i,j \in B_r} 1\left[M_{ij}^k \geq 0.1 \cdot max\, M^k\right]\right\} \quad (26)$$

[0109] 同样，使用积分图的方法来快速计算候选区域的上下文区域在CPG图上的空间密度：

[0110] $$\rho_{rk}^c = \frac{1}{\sqrt{|B_r^c| - |B_r|}}\{\{ii(i_2^c, j_2^c) - ii(i_1^c, j_2^c) - ii(i_2^c, j_1^c) + ii(i_1^c, j_1^c)\} - \{ii(i_2, j_2) - ii(i_1, j_2) - ii(i_2, j_1) + ii(i_1, j_1)\}\} \quad (27)$$

[0111] 实际上，候选区域B_r的上下文区域B_r^c是一个环形的框，也就是原来候选区域B_r的周围空间。因此，上下文区域在CPG图上的空间密度就对应的候选区域周围一圈在CPG图上的空间密度。这和Gidaris等人(Gidaris,Spyros,and Nikos Komodakis."Object detection via a multi-region and semantic segmentation-aware cnn model." Proceedings of the IEEE International Conference on Computer Vision.2015.)提出的multi-regions类似。不同的是Gidaris把多个区域的特征合并成一个特征，然后训练分类器进行分类。本发明是分别计算原来的候选区域和上下文区域在CPG图上的空间密度。把原来候选区域在CPG图上的空间密度减去对应的上下文区域的空间密度得到候选区域的置信度矩阵W：

[0112] $$W_{rk} = \rho_{rk} - \rho_{rk}^c \quad (28)$$

[0113] 其中，$W \in \mathcal{R}^{R \times K}$，并且对每个类别的置信度分别进行归一化，使得W里的每一列里元素最大的值是1:

[0114] $$W_{rk} = \frac{W_{rk}}{\max\limits_{r' < R} W_{r'k}} \quad (29)$$

[0115] 4)在模型训练中，根据置信度矩阵里元素的符号得到两个加权后的得分矩阵：

[0116] $$S^+ = S \odot W \odot 1[W > 0] \quad (30)$$

[0117] $$S^{-} = S \odot W \odot 1[W < 0] \odot (-1) \quad (31)$$

[0118] 其中，⊙表示点对点乘积，S 是原来的候选区域得分矩阵，S^{+}是置信度为正数的候选区域加权后的得分矩阵，S^{-}是置信度为负数的候选区域加权后的得分矩阵，而且 $S, W, S^{+}, S^{-} \in \mathcal{R}^{R \times K}$。最后使用累加池化可以分别得到两个候选区域集合的类别预测结果向量y^{+}和y^{-}，而且 $\boldsymbol{y}^{+}, \ \boldsymbol{y}^{-} \in \mathcal{R}^{K}$。

[0119] $$\boldsymbol{y}_{\boldsymbol{k}}^{+} = \sum_{r=1}^{R} S_{rk}^{+} \quad (32)$$

[0120] $$\boldsymbol{y}_{\boldsymbol{k}}^{-} = \sum_{r=1}^{R} S_{rk}^{-} \quad (33)$$

[0121] 也就是分别把两个候选区域集里所有候选区域的第k个类别的得分累加起来获得对图像第k个类别的预测结果。最后定义交叉熵损失函数为：

[0122] $$L_{label} = \sum_{I} \sum_{k}^{K} \{t_k \log y_k^{+} + (1 - t_k) \log(1 - y_k^{+}) + \log(1 - y_k^{-})\} \quad (34)$$

[0123] 前两项和原来的交叉熵损失函数一样，只不过原来的预测得分y_k是所有的候选区域的预测得分的聚合结果，现在预测得分 y_k^{+}是部分候选区域的预测得分的聚合结果。第三项是用来惩罚置信度为负数的候选区域的预测得分。若置信度为负数的候选区域的预测得分接近于0 的时候，累加池化获得的预测得分 y_k^{-}也接近于0，这时候损失函数的第三项接近于0。

[0124] 5)在模型训练中，本发明还提出了多中心正则化方法。多中心正则化为每个类别维护多个深度卷积神经网络特征中心，定义为：

[0125] $C = \{c_{km} | 0 \leqslant k < K, 0 \leqslant m < M\}$ (35)

[0126] 其中，M表示每个类别的特征中心数目。这里定义图像I中第k个类别的预测得分排名前d 个候选区域的深度卷积神经网络特征集合为：

[0127] $Z(I,k) = \{z_1 \ldots z_d\}$ (36)

[0128] 若图像I里没有第k个类别的目标物体存在，则 $Z(I,k) = \emptyset$。对于每一个特征集合Z(I,k)，定义一个中心选择函数：

[0129] $$m^{*}(I,k) = \arg \min_{0 \leq m < M} \sum_{z \in Z(I,k)} \| z - c_{km} \|_2^2 \quad (37)$$

[0130] 这个函数的含义是对于输入图像I的第k个类别，目标物体的深度卷积神经网络特征中心是第$m^{*}(I,k)$个中心，也就是 $c_{km^{*}(I,k)}$。这里假设每一张图像的每一个存在的类别有且只有一个特征中心。最后定义多中心损失函数为：

[0131] $$L_{center}=\frac{\sum_I\sum_{k=0}^{K}\sum_{z\in Z(I,k)}\|z-c_{km^*(I,k)}\|_2^2}{2\cdot\sum_I\sum_{k=0}^{K}|Z(I,k)|} \quad (38)$$

[0132] 多中心损失函数惩罚图像里每个存在类别的目标物体的深度卷积神经网络特征和对应类别的特征中心的欧式距离。若图像里每个出现的类别的目标物体的深度卷积神经网络特征和对应类别的其中一个特征中心一样，则多中心损失函数为0。训练时候，每个类别的特征中心用高斯分布随机初始化。为了最小化多中心损失，需要计算L_{center}对每个候选区域特征z的导数：

[0133] $$\Delta z=\frac{\sum_I\sum_{k=0}^{K}1[z\in Z(I,k)]\cdot(z-c_{km^*(I,k)})}{\sum_I\sum_{k=0}^{K}|Z(I,k)|} \quad (39)$$

[0134] 由于多中心损失函数L_{center}可以对每个候选区域特征z求导，因此可以把损失误差通过反向传播算法传播到前面神经网络层并影响模型的优化。同时需要计算L_{center}对每个中心c_{km}的导数：

[0135] $$\Delta c_{km}=\frac{\sum_I\sum_{k=0}^{K}1[m^*(I,k)=\mathrm{m}]\cdot(c_{km}-\mathrm{z})}{\sum_I\sum_{k=0}^{K}|Z(I,k)|} \quad (40)$$

[0136] 用以下公式更新每个中心特征：

[0137] $c_{km}:=c_{km}+\sigma\Delta c_{km}$ (41)

[0138] 其中，σ是中心特征的学习速率。

[0139] 6)在模型测试中，只需要把测试图像以及对应的候选区域输入模型，模型输出每个候选区域对于每个类别的预测得分，弱监督目标检测完毕。

[0140] 本发明对聚合的过程进行深入的改进，并分别提出特定类别的像素梯度图(Category-Specific Pixel Gradient map)、候选区域空间约束(Region Spatial Constraint)和多中心正则化(Multi-Center Regularization)等方法改进基于弱监督学习的目标检测算法。本发明的方法在训练的过程中提取特定类别的像素梯度图，特定类别的像素梯度图反应了像素对特定类别的响应，因此模型可以使用特定类别的像素梯度图来粗略估计目标物体的形状和位置，然后根据目标物体的形状和位置的粗略估计结果和候选区域的空间位置的重叠程度计算对应候选区域包含目标物体的置信度。最后把候选区域的置信度引入候选区域分类得分的聚合过程中。因此最后的聚合结果同时包含了候选区域的分类得分和候选区域的空间信息。候选区域的空间约束还能排除大量的背景噪声区域，因此通过学习可以获得更加准确的模型。最后本发明提出使用多中心正则化来保证模型的学习过程更加的稳定。

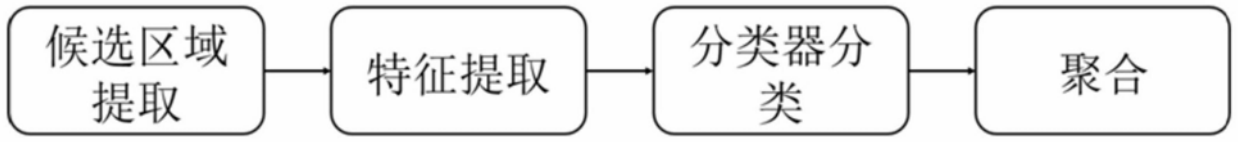
候选区域提取
特征提取
分类器分类
聚合

图1

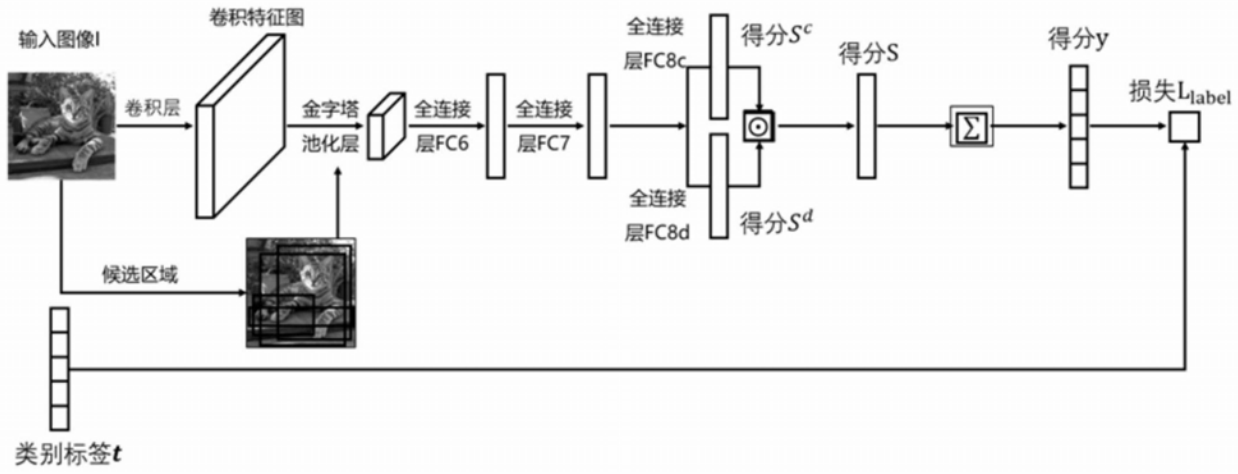
输入图像I
卷积特征图
卷积层
金字塔池化层
全连接层FC6
全连接层FC7
全连接层FC8c
全连接层FC8d
得分S^c
得分S^d
得分S
得分y
损失L_{label}
候选区域
类别标签t

图2

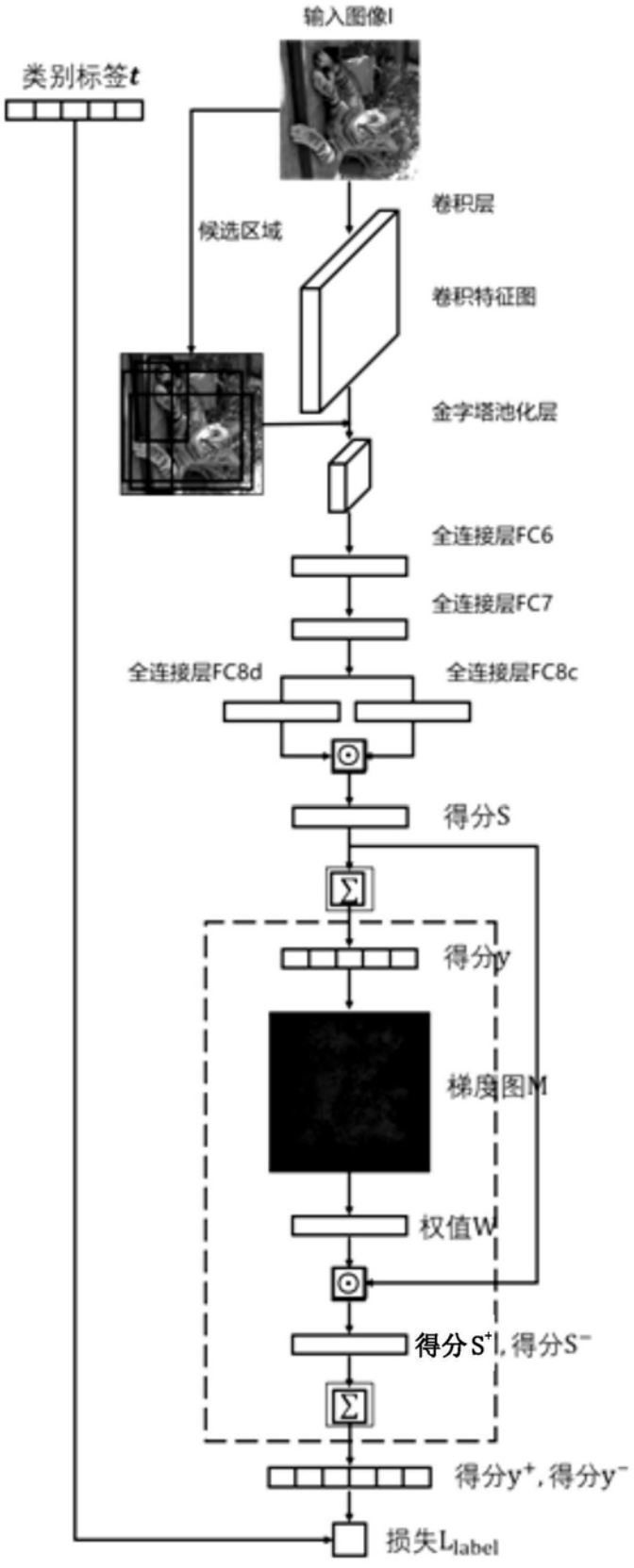
输入图像I
类别标签t
候选区域
卷积层
卷积特征图
金字塔池化层
全连接层FC6
全连接层FC7
全连接层FC8d
全连接层FC8c
得分S
得分y
梯度图M
权值W
得分S^+,得分S^-
得分y^+,得分y^-
损失L_{label}

图3

图4