



# (12)发明专利

(10)授权公告号 CN 106683091 B

(45)授权公告日 2019.09.24

(21)申请号 201710010558.1

CN 104992142 A, 2015.10.21,

(22)申请日 2017.01.06

CN 105447473 A, 2016.03.30,

(65)同一申请的已公布的文献号

CN 101533467 A, 2009.09.16,

申请公布号 CN 106683091 A

CN 104732243 A, 2015.06.24,

US 2016358038 A1, 2016.12.08,

(43)申请公布日 2017.05.17

审查员 刘海艳

(73)专利权人 北京理工大学

地址 100081 北京市海淀区中关村南大街5号

(72)发明人 刘明 杜浩源 董立泉 赵跃进

刘小华 惠梅 孔令琴

(51)Int.Cl.

G06T 7/00(2017.01)

G06N 3/04(2006.01)

(56)对比文件

CN 101533467 A, 2009.09.16,

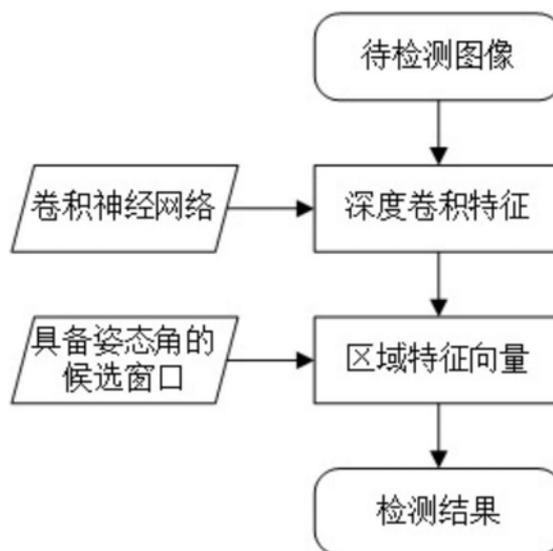
权利要求书1页 说明书4页 附图4页

(54)发明名称

一种基于深度卷积神经网络的目标分类及姿态检测方法

(57)摘要

本发明属于图像处理与目标检测领域,公开了一种基于深度卷积神经网络的目标分类及姿态检测方法。所述方法核心内容包括:1.样本集的制作方法;2.候选窗口生成方法;3.特征提取方法;4.网络训练方法;5.目标标注方法。待检测图像经过卷积神经网络得到深度卷积特征后,将具备姿态角的候选窗口映射到该特征层上,得到方向性的区域特征向量,通过该特征向量进行分类和预测得到最终的检测结果。该方法能够从样本中提取更加纯净的目标特征,可提高分类的正确率,并且能够实现目标姿态角检测功能。



1. 一种基于深度卷积神经网络的目标分类及姿态检测方法,其特征在于包括以下步骤:

a) 制作样本集,包括训练集和测试集,每张样本图像可包含多个目标,每个目标用六个参数  $(l, x, y, h, w, \theta)$  进行表示,  $l$  代表该目标的标签,  $x$  和  $y$  代表目标中心在图像中的横纵坐标,  $h$  和  $w$  代表目标真值窗口的高和宽,  $\theta$  表示目标相对于基准轴的姿态角;

b) 候选窗口生成过程,在整幅图像上以一定步长均匀生成候选窗口,窗口采用不同尺寸和不同姿态角的矩形框;

c) 深度特征提取过程,先将待检测的图像输入到预先训练的模型进行计算得到深度卷积特征图,将b) 当中生成的候选窗口依据图像金字塔的机制映射到深度卷积特征图上面生成固定维度的特征向量;

d) 子网络构建过程,在获得的卷积特征层的基础上,构建三个子网络分别用于识别目标种类、修正候选窗口位置和预测目标姿态角,对于网络中新添加的层,采用Xavier初始化方法建立参数;

e) 训练样本的生成过程,样本中不仅包含图像,还包含候选窗口的正负样本,正负样本依据候选窗口与真值窗口的重叠比例IOU进行划分,使样本集中正负样本比例接近于1:1;

f) 网络训练过程,通过批量随机梯度下降法对网络参数进行调整,每次训练需至少输入一幅图像,保证正样本数量的前提下随机抽取负样本,在训练起始阶段通过较大学习率对参数进行更新,在网络输出误差减小缓慢时减小学习率;

g) 目标标注步骤,输入一幅待检测图像,利用训练完成的网络进行计算,得到候选窗口的类别、位置偏移量、姿态角和分数,设定阈值排除分数较低的窗口,通过极大值抑制方法保留分数较高的窗口作为最终检测结果,利用具备姿态角的矩形窗口框选目标位置。

2. 根据权利要求1所述的方法,其特征在于,步骤a中所述姿态角通过对目标构成点阵做最小面积包围矩形得到,该包围矩形长轴方向与基准轴之间的夹角即为姿态角。

3. 根据权利要求1所述的方法,其特征在于,步骤b中所述候选窗口通过正立矩形框进行旋转得到,旋转中心为在图像上以一定步长均匀分布的参考点。

4. 根据权利要求1所述的方法,其特征在于,步骤f中所述训练方法包括:

采用批量梯度下降对网络参数进行迭代更新,网络的代价函数由多项构成,公式为

$$J = \frac{1}{N} [J_1(l'_i, l_i) + \lambda \cdot 1\{l_i \geq 1\} \cdot J_2(p'_i, p_i) + \mu \cdot 1\{\theta_i \geq 1\} \cdot J_3(\theta'_i, \theta_i)],$$

式中,  $N$  表示单次批量随机梯度下降采用的样本数量;  $J_1$  为分类误差项,代表候选窗口的分类误差,  $l'_i$  为预测标签,  $l_i$  为目标真实标签;  $J_2$  为位置误差项,表示候选窗口与真值窗口的位置偏差,  $p'_i$  是由网络输出的对候选窗口位置的修正量,  $p_i$  是真值相对于候选窗口的位置偏差量,  $1\{\cdot\}$  为指示函数,当花括号内表述为真时等于1,否则等于0;  $J_3$  为角度误差项,用于计算候选窗口与真值窗口的角度误差,  $\theta'_i$  为网络输出的目标姿态角对候选窗口角度的修正量,  $\theta_i$  为目标真实姿态角与候选窗口角度的偏差量;  $\lambda$  和  $\mu$  用来平衡三项之间的权重。

## 一种基于深度卷积神经网络的目标分类及姿态检测方法

### 技术领域

[0001] 本发明属于图像处理与目标检测领域,公开了一种基于深度卷积神经网络的目标分类及姿态检测方法。

### 背景技术

[0002] 目标检测需要在给定的图像上精确地标记目标所在位置,并识别出物体的种类。目标尺寸大小不定、目标在图像当中的位置不定、目标的姿态角度不定以及背景光照变化等将会对检测过程造成困难,导致检测正确率下降。

[0003] 经典的目标检测方法通常采用模板匹配加滑动窗口的方式来解决目标在图像中的识别和定位问题,这种方法耗时长且在目标变化剧烈的情况下效果较差。卷积神经网络的出现带动了这一课题的发展,大幅提高了检测精度(mAP)。

[0004] 纽约大学的Pierre Sermanet等人提出的Overfeat网络采用多尺度的滑动窗口在同一卷积特征图上进行目标定位和识别,但该方法效果和速度都不尽如人意。Ross Girshick等人提出的RCNN方法基于选择性搜索方法(Selective search)预测物体可能存在的区域,但存在着仿射变换破坏候选区域特征的问题。SPP-NET方法和Fast RCNN方法仍然通过选择性搜索的方式生成候选区域,利用图像金字塔机制解决了候选区域的尺度映射问题,同时实现了多阶段连续训练。Faster RCNN用一个单独的区域生成网络代替了选择性搜索的方法,解决了候选区域生成过程耗费大量时间的问题,加快了网络的训练速度和测试速度,也提高了检测结果的准确性。

[0005] 根据上述每种方法的发明者撰写的学术论文内容显示,目前目标检测方法通常所采用的选框为与图像的四条边框分别平行的正立矩形,这种形式的选框仅对处于正立姿态角的目标(目标的几何中心线与图像边框近似平行或垂直)具有较好的检测正确率,从另一角度解释则为当前检测方法没有考虑到目标姿态角的变化。在实际检测情景当中,目标经常处于倾斜姿态,即中心线与图像边框存在一定幅度的夹角。因此若采用正立矩形框所框选的目标中会包含很多背景信息,造成信噪比大幅下降,不利于后续的目标分类和位置校正。并且由于方法并未考虑到目标的在图像中的姿态角变化,因此无法表征目标的姿态。综合上述观点,当前目标检测方法具备的检测能力有限,并且对于目标状态的表征不够充分。

### 发明内容

[0006] 本发明提出一种目标检测方法。主要解决的问题识别目标的类别,定位目标在图像中的位置、尺寸和姿态角。

[0007] 本发明是一种基于深度卷积神经网络的目标分类及姿态检测方法,包括样本集制作方法,神经网络的训练方法和目标的标注方法。

[0008] 所述样本集的制作方法扩充了传统样本集的内容,传统的样本集中只包含目标种类和目标左上角点、右下角点的横纵坐标,然而在目标倾斜的情形下,这种标注方式并不能准确的反应目标姿态。在本发明中,除记录目标的种类外,还记录目标中心点在图像中的

横纵坐标,采用面积最小的矩形窗口框选目标,将矩形窗口的长边和短边所占像素数作为目标的宽和高,将矩形窗口长轴方向与基准轴之间的夹角作为目标的姿态角。

[0009] 所述卷积神经网络的训练方法包括如下步骤:

[0010] 输入批量训练图像 $S_i, i=1, \dots, M$ ,  $M$ 代表批量训练图像的数量,同时需要输入图像中目标的标签及相对应的位置,对训练图像的尺寸进行归一化;

[0011] 将图像划分为不同尺寸和不同姿态角的候选窗口,计算出所有候选窗口与真值窗口的重叠区域比例,通过设定正负阈值,将候选窗口分类为正负样本;

[0012] 从每幅图像当中选取 $N/M$ 个候选窗口作为训练样本, $N$ 代表批量随机梯度下降采用的样本数量,控制正负样本数量的比例接近于1:1,每个样本 $a_j$ 均包含六个参数 $(l_j, x_j, y_j, h_j, w_j, \theta_j)$ ,  $j=1, \dots, N$ ,  $l_j$ 代表该样本的标签,  $x_j, y_j$ 代表候选窗口中心在图像中的横纵坐标,  $h_j, w_j$ 代表候选窗口的高和宽,  $\theta_j$ 表示候选窗口倾斜的角度。

[0013] 通过网络前向传播提取样本的深度卷积特征,将候选窗口映射到该特征上,得到固定维度的特征向量;

[0014] 在特征向量层上建立三个独立的网络,分别利用特征向量计算候选窗口的分类误差、位置偏差和角度误差;

[0015] 网络输出的代价函数为:

$$[0016] \quad J(W, b) = \frac{1}{N} \left[ J_1(l_i', l_i) + \lambda \cdot 1\{l_i \geq 1\} \cdot J_2(p_i', p_i) + \mu \cdot 1\{\theta_i \geq 1\} \cdot J_3(\theta_i', \theta_i) \right],$$

[0017] 式中, $W$ 和 $b$ 为网络需要调整的参数; $N$ 表示单次批量随机梯度下降采用的样本数量; $J_1$ 为分类误差项,代表候选窗口的分类误差,  $l_i'$ 为预测标签,  $l_i$ 为目标真实标签; $J_2$ 为位置误差项,表示候选窗口与真值窗口的位置偏差,  $p_i'$ 是由网络输出的对候选窗口位置的修正量,  $p_i$ 是真值相对于候选窗口的位置偏差量,  $1\{\cdot\}$ 为指示函数,当花括号内表述为真(样本为正样本)时等于1,否则等于0; $J_3$ 为角度误差项,用于计算候选窗口与真值窗口的角度误差,  $\theta_i'$ 为网络输出的目标姿态角对候选窗口角度的修正量,  $\theta_i$ 为目标真实姿态角与候选窗口角度的偏差量; $\lambda$ 和 $\mu$ 用来平衡三项之间的权重。

[0018] 通过批量随机梯度下降的反向传播方法对网络各层参数进行调整,当迭代次数达到预设的最大值或者网络输出的误差收敛时即停止训练。

[0019] 所述目标位置的标注方法包括如下步骤:

[0020] 第一步,输入待检测图像,将图像的大小归一化为网络的固定输入尺寸;

[0021] 第二步,针对每个候选窗口,通过网络计算得出目标类别、位置修正量、姿态角和分数;

[0022] 第三步,设定目标尺寸阈值和分数阈值,保留一定数量的候选窗口;

[0023] 第四步,通过极大值抑制方法,进一步去除重叠率较高且分数较低的候选窗口;

[0024] 第五步,根据结果在图像中以具备姿态角的矩形窗口的框选出目标范围。

[0025] 有益效果

[0026] 本发明在图像中以固定间隔划分不同尺寸和不同姿态角的候选窗口,适于对不同大小和不同姿态角的目标进行检测,避免了选择性搜索方法的大量计算。相比于传统的矩形窗口,具备姿态角的窗口可以更加准确地框选处于倾斜姿态的物体,减少了窗口中包含的背景信息,通过图像金字塔的映射机制将倾斜窗口映射至深度卷积特征层上,可以提取

到更纯净的特征,解决了现有目标分类方法对姿态敏感的问题,有效地提高了目标检测的准确率。此外,本方法还具备目标姿态角检测功能,对于图像中目标状态的表征更加完备。

### 附图说明

- [0027] 图1是依照本发明实施例的目标检测框架;  
 [0028] 图2是依照本发明实施例将姿态角进行离散化的示意图;  
 [0029] 图3是依照本发明实施例的网络训练流程;  
 [0030] 图4是依照本发明实施例的特征提取网络和目标检测网络;  
 [0031] 图5是依照本发明实施例对目标进行检测并标注的结果示意图。

### 具体实施方式

[0032] 为了更加清楚地表述本发明的目的、技术方案和优点,以下结合具体实施例,并参照附图,给出了本发明进一步详细说明,但本发明的保护范围不限于下述的实施例。

[0033] 本实施例当中对目标进行检测的框架如图1所示,图像经过卷积神经网络得到深度卷积特征后,将具备姿态角的候选窗口映射到该特征层上,得到方向性的区域特征向量,再通过对特征向量进行分类和预测得到最终的检测结果。

[0034] 本实施例采用的数据集部分源自公开图像库PASCAL VOC 2007和PASCAL VOC 2012,此外我们还添加了许多人工采集到的不同场景和不同角度下目标图像。最后得到的数据集中包含20类目标的彩色图像,共计15000张,其中10000张用于训练,5000张用于测试。数据集的标注当中包含目标的种类、目标的中心坐标和宽高像素数、目标的姿态角。

[0035] 为降低网络对姿态角检测的难度,我们将姿态角进行离散化,使角度检测成为分类任务而不是回归任务。角度的离散如图2所示,将360度的方位角平分为16个区间,考虑到目标识别须具备旋转对称性,将对顶角的角度区间视为同一角度类别。以x轴正半轴为0°方位角、逆时针方向定义角度值,角度类别与角度区间对照如表1所示。

[0036] 表1

[0037]

|      |                                   |                                    |                                    |                                    |
|------|-----------------------------------|------------------------------------|------------------------------------|------------------------------------|
| 角度类别 | I                                 | II                                 | III                                | IV                                 |
| 角度区间 | 348.75°~11.25°<br>168.75°~191.25° | 11.25°~33.75°<br>191.25°~213.75°   | 33.75°~56.25°<br>213.75°~236.25°   | 56.25°~78.75°<br>236.25°~258.75°   |
| 角度类别 | V                                 | VI                                 | VII                                | VIII                               |
| 角度区间 | 78.75°~101.25°<br>258.75°~281.25° | 101.25°~123.75°<br>281.25°~303.75° | 123.75°~146.25°<br>303.75°~326.25° | 146.25°~168.75°<br>326.25°~348.75° |

[0038] 训练的模型由特征提取网络(图-3 左)和目标检测网络(图-3 右)两部分构成,特征提取网络选取在ILSVRC2015数据库上预训练的VGG-16网络模型中的一部分,去除掉VGG-16网络最后一个最大池化层后面的层,保留下来的网络包含13个卷积层和4个最大池化层,在此基础上添加一层ROI pooling输出区域特征向量。目标检测网络由两个公共的全连接层、两个公共的dropout层和三个分支的全连接层构成,三个分支分别得出目标的类别、候选窗口的修正量和目标姿态角。

[0039] 倾斜候选窗口的生成过程为:根据特征提取网络中卷积和池化操作的步长,在原图像上以16个像素为间隔均匀选取采样点,以采样点为中心生成四种尺寸(32, 64, 128,

256)的基准框,在每个基准框中参照表1划分出8个姿态角的候选窗口,也就是在每个采样点周围生成共计32个倾斜候选窗口。

[0040] 网络的训练过程如图4所示,描述如下:

[0041] 输入训练图像和图像中所有目标的类别及相对应的位置信息。将图像尺寸归一化至固定的网络输入大小,图像像素值通过减去平均值做零均值处理;

[0042] 计算出所有倾斜候选窗口与真值窗口的重叠区域比例,当比例大于0.7时,认为候选窗口为正样本,当比例小于0.3时,则认为是负样本;

[0043] 训练方法采用批量梯度下降,本实施例中,每批次选取2幅图像,每幅图像上随机采样128个样本,使正负样本数量的比例接近于1:1;

[0044] 通过网络前向传播提取样本的深度卷积特征,将候选窗口映射到该特征上,通过ROI pooling得到大小为 $7 \times 7$ 的区域特征向量;

[0045] 在特征向量层上建立三个独立的网络,分别利用特征向量计算候选窗口的分类误差、位置偏差和角度分类误差;

[0046] 网络输出的代价函数为:

$$[0047] \quad J(W, b) = \frac{1}{N} [J_1(l'_i, l_i) + \lambda \cdot 1\{l_i \geq 1\} \cdot J_2(p'_i, p_i) + \mu \cdot 1\{\theta_i \geq 1\} \cdot J_3(\theta'_i, \theta_i)],$$

[0048] 式中,W和b为网络需要调整的参数;N表示单次批量随机梯度下降采用的样本数量; $J_1$ 为分类误差项,代表候选窗口的分类误差, $l_i'$ 为预测标签, $l_i$ 为目标真实标签; $J_2$ 为位置误差项,表示候选窗口与真值窗口的位置偏差, $p_i'$ 是由网络输出的对候选窗口位置的修正量, $p_i$ 是真值相对于候选窗口的位置偏差量, $1\{\cdot\}$ 为指示函数,当花括号内表述为真(样本为正样本)时等于1,否则等于0; $J_3$ 为角度误差项,用于计算候选窗口与真值窗口的角度误差, $\theta_i'$ 为网络输出的目标姿态角对候选窗口角度的修正量, $\theta_i$ 为目标真实姿态角与候选窗口角度的偏差量; $\lambda$ 和 $\mu$ 用来平衡三项之间的权重。

[0049] 通过批量随机梯度下降的反向传播方法对网络各层参数进行调整,当迭代次数达到预设的最大值或者网络输出的误差收敛时即停止训练。

[0050] 对目标进行检测并标注大致流程为:输入待检测图像,将图像的大小归一化为网络的固定输入尺寸。针对每个候选窗口,通过网络计算得出目标类别、位置修正量、姿态角和分数。设定目标尺寸阈值和分数阈值,保留一定数量的候选窗口。通过极大值抑制方法,进一步去除重叠率较高且分数较低的候选窗口。根据结果在图像中以倾斜矩形框的方式框选出目标范围,结果如图5所示。

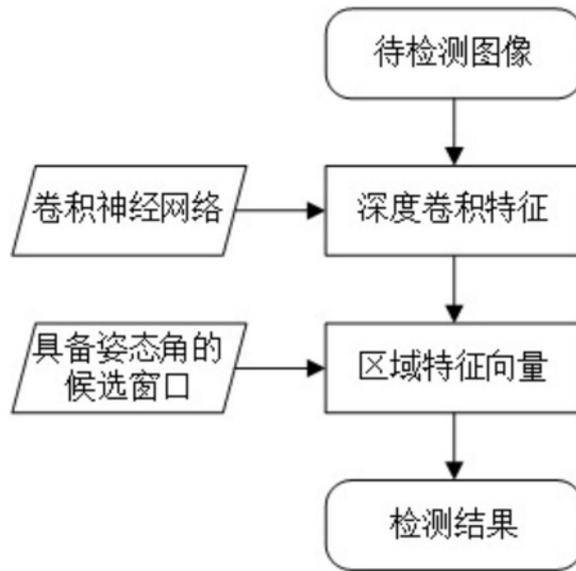


图1

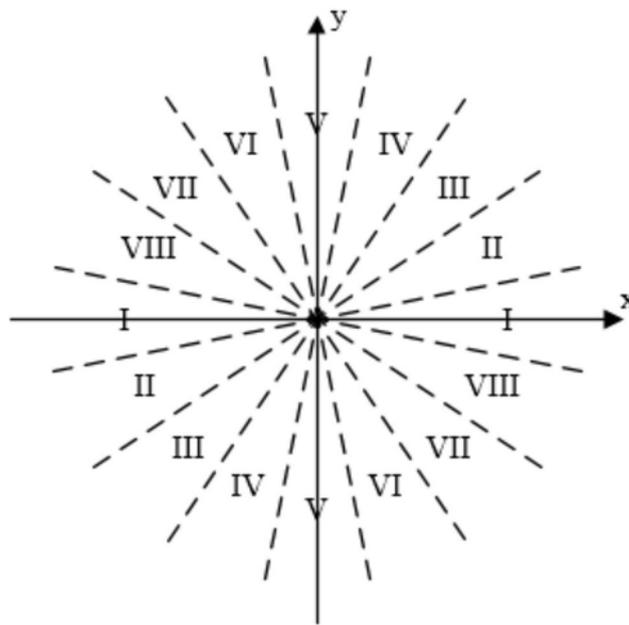


图2

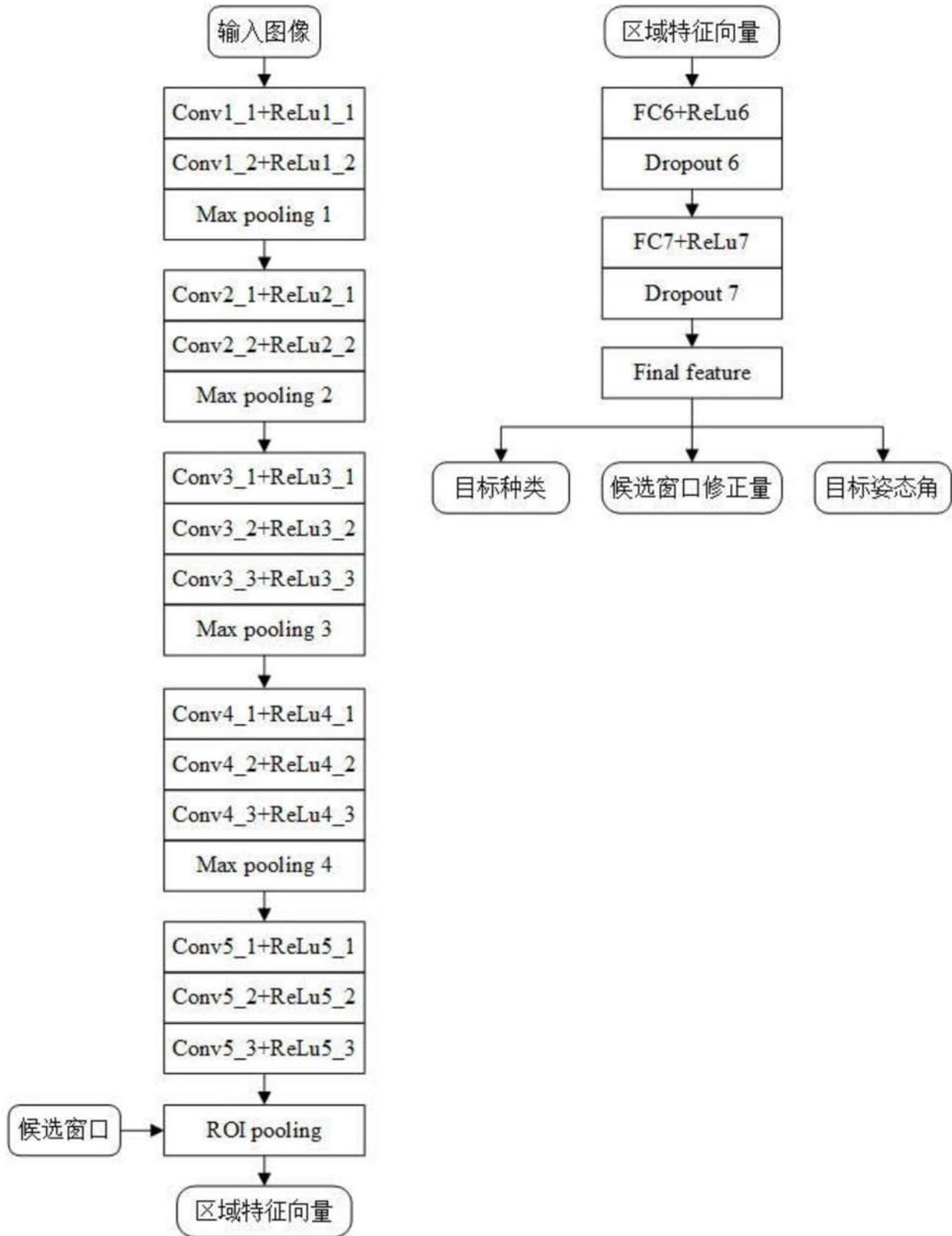


图3

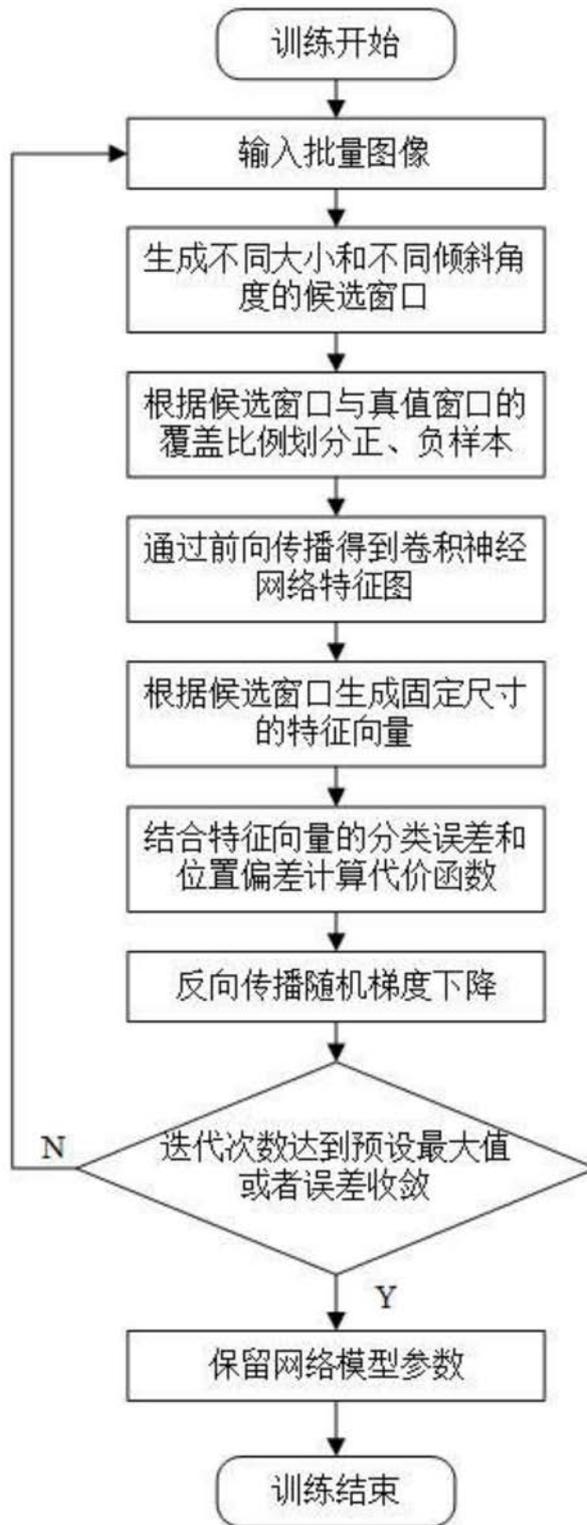


图4



图5