



(12) 发明专利申请

(10) 申请公布号 CN 113313118 A

(43) 申请公布日 2021.08.27

(21) 申请号 202110712902.8

G06N 3/08 (2006.01)

(22) 申请日 2021.06.25

(71) 申请人 哈尔滨工程大学

地址 150001 黑龙江省哈尔滨市南岗区南通大街145号

(72) 发明人 宿南 李瑞 王金鹏 闫奕名
赵春晖 黄博闻

(74) 专利代理机构 哈尔滨市松花江专利商标事务所 23109

代理人 岳泉清

(51) Int.Cl.

G06K 9/32 (2006.01)

G06K 9/46 (2006.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

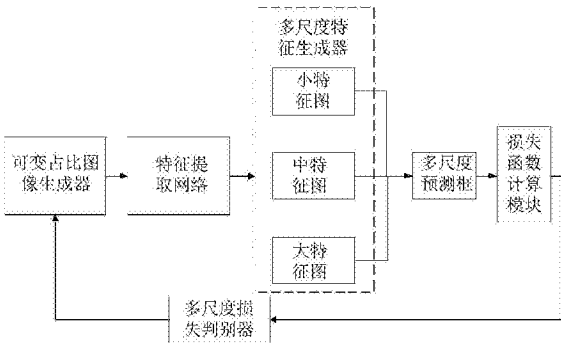
权利要求书2页 说明书6页 附图2页

(54) 发明名称

基于多尺度特征融合的自适应可变占比目标检测方法

(57) 摘要

基于多尺度特征融合的自适应可变占比目标检测方法,属于目标检测技术领域,本发明为解决待测目标是小目标情况下,小目标在特征图上尺寸过小,学习过程中易被大目标主导,导致小目标检测效果较差的问题。它包括:可变占比图像生成器将原始数据生成不同大小的目标数据,然后将目标数据输送至特征提取网络,特征提取网络对特征进行提取,然后将提取的特征输送至多尺度特征生成器,多尺度特征生成器产生三个尺度的特征图,所述三个尺度特征图包括小特征图、中特征图和大特征图,将三个尺度的特征图输送至多尺度预测框中,多尺度预测框对三个尺度的特征图分别生成预测框,预测框内即为目标检测结果。本发明用于目标检测,尤其适用于小目标的目标检测。



句式可用

句式可用

1. 基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 它基于Yolo算法框架实现; 该目标检测方法包括:

可变占比图像生成器、特征提取网络、多尺度特征生成器和多尺度预测框;

可变占比图像生成器将原始数据生成不同大小的目标数据, 然后将不同大小的目标数据输送至特征提取网络, 特征提取网络对特征进行提取, 然后将提取的特征输送至多尺度特征生成器, 多尺度特征生成器产生三个尺度的特征图, 所述三个尺度特征图包括小特征图、中特征图和大特征图, 将三个尺度的特征图输送至多尺度预测框中, 多尺度预测框对三个尺度的特征图分别生成预测框, 预测框内即为目标检测结果。

2. 根据权利要求1所述的基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 可变占比图像生成器输入的原始数据为:

$$F_{\text{imgin}}(x, y) \mid_{0 < x, y < t};$$

将原始数据平移得到: $F_{\text{imgin}}(x, y) \mid_{t1 < x, y < t+t1};$

其中, $t1$ 表示填充的黑色边带的宽度, 受多尺度损失反馈调节, 填充后的图片函数为:

$$F_{\text{imgout}}(x, y) \mid_{v=0 < x, y < t+2*t1} = F_{\text{imgin}}(x, y) \mid_{\omega=t1 < x, y < t+t1} + (0, 0, 0) \mid_{v-\omega};$$

其中, $(0, 0, 0) \mid_{\alpha}$ 为在 α 区域填充黑色; α 表示以距离原图像 $t1$ 的直线为边界的内侧区域, ω 表示填充前的定义域边界, v 表示填充后的定义域边界,

输出的目标数据为:

$$F_{\text{out}}(x, y) \mid_{0 < x, y < t} = F_{\text{imgout}}(x / [(t+2*t1) / t], y / [(t+2*t1) / t]) \mid_{v=0 < x < t+2*t1};$$

(x, y) 表示二维平面上点的位置。

3. 根据权利要求2所述的基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t+2*t1}$ 倍。

4. 根据权利要求3所述的基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 该目标检测方法还包括: 损失函数计算模块和多尺度损失判别器;

损失函数计算模块分别对三个尺度的特征图计算损失函数, 多尺度损失判别器根据损失函数的大小调节可变占比图像生成器的填充区域, 当小特征图无法获得目标检测结果时, 增大 $t1$, 使得可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t+2*t1}$ 倍。

5. 根据权利要求2所述的基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 特征提取网络包括五个残差单元;

可变占比图像生成器的输出数据 $F_{\text{out}}(x, y) \mid_{0 < x, y < t}$ 经过五个残差单元的五次下采样运算获取小特征图预测所需的特征图 $f1$;

$f1$ 经过多尺度池化层和上采样, 然后与中浅特征层四次下采样的结果进行拼接, 获得中特征图预测所需的特征图 $f2$;

$f2$ 经过多尺度池化层和上采样, 然后与浅特征层四次下采样的结果进行拼接, 获得大特征图预测所需的特征图 $f3$ 。

6. 根据权利要求5所述的基于多尺度特征融合的自适应可变占比目标检测方法, 其特征在于, 所述多尺度特征生成器产生三个尺度特征图的具体方法包括:

将提取的特征经过多尺度特征池化模块,捕捉不同尺度的特征,然后将不同尺度的特征经过多尺度特征融合;

多尺度特征包括:

将大特征图预测所需的特征图f3的特征经过下采样后与中特征图预测所需的特征图f2进行融合,获得中特征图,

将中特征图预测所需的特征图f2的特征经过下采样后与小特征图预测所需的特征图f1进行融合,获得小特征图,

大特征图预测所需的特征图f3的特征不变,获得大特征图。

7.根据权利要求6所述的基于多尺度特征融合的自适应可变占比目标检测方法,其特征在于,所述小特征图的大小为13*13,中特征图的大小为26*26,大特征图的大小为52*52。

8.根据权利要求6所述的基于多尺度特征融合的自适应可变占比目标检测方法,其特征在于,所述多尺度特征池化模块包括四个不同大小的池化结构,能够捕捉四个不同尺度的特征,四个不同大小池化结构的最大池化结构max_pooli表示*i***i*最大池化,此处*i*=1,5,9,13。

9.根据权利要求8所述的基于多尺度特征融合的自适应可变占比目标检测方法,其特征在于,所述多尺度特征池化模块的函数表达式为:

$$F_{MSP_out} = \text{concat}(\text{max_pool1}(F_{MSP_in}), \text{max_pool5}(F_{MSP_in}), \text{max_pool9}(F_{MSP_in}), \text{max_pool13}(F_{MSP_in}))$$

其中, F_{MSP_in} 表示多尺度特征池化模块的输入特征, F_{MSP_out} 表示多尺度特征池化模块的输出特征,concat表示对特征张量进行拼接。

10.根据权利要求6所述的基于多尺度特征融合的自适应可变占比目标检测方法,其特征在于,多尺度预测框对三个尺度的特征图分别生成预测框的具体方法包括:

多尺度预测框在每个尺度的特征图的每个网格上设置三个先验框,获得预测值,每个预测值是(4+1+n)维向量,4表示位置信息,1表示边框置信度,n表示n类目标的概率,n类目标是数据集中包含的目标类别;

对预测值与人工标注框计算交并比,保留交并比大于阈值的预测值,如交并比均小于阈值,则保留边框置信度排序为前三位的预测值;

对保留的预测值进行非极大值抑制,去除重叠大于阈值的边框,获得预测框。

增加对
特征图
尺寸的描述

基于多尺度特征融合的自适应可变占比目标检测方法

技术领域

[0001] 本发明涉及一种自适应可变占比目标检测方法,属于目标检测技术领域。

背景技术

[0002] 目标检测也称为目标提取,是在图像中定位并分类出感兴趣的目標的方法。在实际应用环境中,由于相机成像存在远小近大的特点,因此,当出现场景较为复杂、待检测目标尺度大小差异较大、待检测目标为小目标等问题时,现有的算法很难满足实际需要。

[0003] 针对待检测目标为小目标的情况,由于小目标在图像中尺寸比较小,在通用目标检测模型中,一般的基础骨干神经网络都有下采样过程,导致原本较小的目标在特征图上的尺寸更加小,这会导致设计的分类器的效果较差,另外由于小目标特征的感受野映射回原图将可能远远大于实际尺寸,造成检测效果差。此外,由于实际的数据中可能存在小目标样本较少,导致神经网络在学习的过程中被大目标主导,导致小目标的检测效果很差。

修改
可用

发明内容

[0004] 本发明目的是为了解决待测目标是小目标的情况下,小目标在特征图上的尺寸过小,以及学习过程中易被大目标主导,导致小目标的检测效果较差的问题,提供了一种基于多尺度特征融合的自适应可变占比目标检测方法。

修改
可用

[0005] 本发明所述基于多尺度特征融合的自适应可变占比目标检测方法,它基于Yolo算法框架实现;该目标检测方法包括:

[0006] 可变占比图像生成器、特征提取网络、多尺度特征生成器和多尺度预测框;

[0007] 可变占比图像生成器将原始数据生成不同大小的目标数据,然后将不同大小的目标数据输送至特征提取网络,特征提取网络对特征进行提取,然后将提取的特征输送至多尺度特征生成器,多尺度特征生成器产生三个尺度的特征图,所述三个尺度特征图包括小特征图、中特征图和大特征图,将三个尺度的特征图输送至多尺度预测框中,多尺度预测框对三个尺度的特征图分别生成预测框,预测框内即为目标检测结果。

[0008] 优选的,可变占比图像生成器输入的原始数据为:

[0009] $F_{\text{imgin}}(x, y) \mid_{0 < x, y < t}$;

[0010] 将原始数据平移得到: $F_{\text{imgin}}(x, y) \mid_{t1 < x, y < t+t1}$;

[0011] 其中, $t1$ 表示填充的黑色边带的宽度,受多尺度损失反馈调节,填充后的图片函数为:

[0012] $F_{\text{imgout}}(x, y) \mid_{v=0 < x, y < t+2*t1} = F_{\text{imgin}}(x, y) \mid_{\omega=t1 < x, y < t+t1} + (0, 0, 0) \mid_{v-\omega}$;

[0013] 其中, $(0, 0, 0) \mid_a$ 为在 a 区域填充黑色; a 表示以距离原图像 $t1$ 的直线为边界的内侧区域, ω 表示填充前的定义域边界, v 表示填充后的定义域边界,

[0014] 输出的目标数据为:

[0015] $F_{\text{out}}(x, y) \mid_{0 < x, y < t} = F_{\text{imgout}}(x/[(t+2*t1)/t], y/[(t+2*t1)/t]) \mid_{v=0 < x < t+2*t1}$;

[0016] (x, y) 表示二维平面上点的位置。

[0017] 优选的,可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t+2*t1}$ 倍。

[0018] 优选的,该目标检测方法还包括:损失函数计算模块和多尺度损失判别器;

[0019] 损失函数计算模块分别对三个尺度的特征图计算损失函数,多尺度损失判别器根据损失函数的大小调节可变占比图像生成器的填充区域,当小特征图无法获得目标检测结果时,增大t1,使得可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t+2*t1}$ 倍。

[0020] 优选的,特征提取网络包括五个残差单元;

[0021] 可变占比图像生成器的输出数据 $F_{out}(x,y)_{0<x,y<t}$ 经过五个残差单元的五次下采样运算获取小特征图预测所需的特征图f1;

[0022] f1经过多尺度池化层和上采样,然后与中浅特征层四次下采样的结果进行拼接,获得中特征图预测所需的特征图f2;

[0023] f2经过多尺度池化层和上采样,然后与浅特征层四次下采样的结果进行拼接,获得大特征图预测所需的特征图f3。

[0024] 优选的,所述多尺度特征生成器产生三个尺度特征图的具体方法包括:

[0025] 将提取的特征经过多尺度特征池化模块,捕捉不同尺度的特征,然后将不同尺度的特征经过多尺度特征融合;

[0026] 多尺度特征包括:

[0027] 将大特征图预测所需的特征图f3的特征经过下采样后与中特征图预测所需的特征图f2进行融合,获得中特征图,

[0028] 将中特征图预测所需的特征图f2的特征经过下采样后与小特征图预测所需的特征图f1进行融合,获得小特征图,

[0029] 大特征图预测所需的特征图f3的特征不变,获得大特征图。

[0030] 优选的,所述小特征图的大小为13*13,中特征图的大小为26*26,大特征图的大小为52*52。

[0031] 优选的,所述多尺度特征池化模块包括四个不同大小的池化结构,能够捕捉四个不同尺度的特征,四个不同大小池化结构的最大池化结构max_pooli表示i*i最大池化,此处i=1,5,9,13。

[0032] 优选的,所述多尺度特征池化模块的函数表达式为:

[0033] $F_{MSP_out} = \text{concat}(\text{max_pool1}(F_{MSP_in}), \text{max_pool5}(F_{MSP_in}),$

[0034] $\text{max_pool9}(F_{MSP_in}), \text{max_pool13}(F_{MSP_in}))$

[0035] 其中, F_{MSP_in} 表示多尺度特征池化模块的输入特征, F_{MSP_out} 表示多尺度特征池化模块的输出特征,concat表示对特征张量进行拼接。

[0036] 优选的,多尺度预测框对三个尺度的特征图分别生成预测框的具体方法包括:

[0037] 多尺度预测框在每个尺度的特征图的每个网格上设置三个先验框,获得预测值,每个预测值是(4+1+n)维向量,4表示位置信息,1表示边框置信度,n表示n类目标的概率,n类目标是数据集中包含的目标类别;

[0038] 对预测值与人工标注框计算交并比,保留交并比大于阈值的预测值,如交并比均小于阈值,则保留边框置信度排序为前三位的预测值;

[0039] 对保留的预测值进行非极大值抑制,去除重叠大于阈值的边框,获得预测框。本发明的优点:本发明提出的基于多尺度特征融合的自适应可变占比目标检测方法利用对三个尺度的损失的大小来反馈调节图像可变占比算法,从而能够改善算法对检测效果较差尺度的检测能力,而且由于会产生小目标样本,可以提高小目标的检测精度,能够提高算法的对不同大小目标的鲁棒性。

附图说明

[0040] 图1是本发明所述基于多尺度特征融合的自适应可变占比目标检测方法的原理框图;

[0041] 图2是本发明所述多尺度特征池化模块的原理框图;

[0042] 图3是本发明所述多尺度特征生成器产生三个尺度特征图的原理框图。

具体实施方式

[0043] 下面将结合本发明实施例中的附图,对本发明实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅仅是本发明一部分实施例,而不是全部的实施例。基于本发明中的实施例,本领域普通技术人员在没有作出创造性劳动的前提下所获得的所有其他实施例,都属于本发明保护的范围。

[0044] 需要说明的是,在不冲突的情况下,本发明中的实施例及实施例中的特征可以相互组合。

[0045] 下面结合附图和具体实施例对本发明作进一步说明,但不作为本发明的限定。

[0046] 具体实施方式一:下面结合图1说明本实施方式,本实施方式所述基于多尺度特征融合的自适应可变占比目标检测方法,它基于Yolo算法框架实现;该目标检测方法包括:

[0047] 可变占比图像生成器、特征提取网络、多尺度特征生成器和多尺度预测框;

[0048] 可变占比图像生成器将原始数据生成不同大小的目标数据,然后将不同大小的目标数据输送至特征提取网络,特征提取网络对特征进行提取,然后将提取的特征输送至多尺度特征生成器,多尺度特征生成器产生三个尺度的特征图,所述三个尺度特征图包括小特征图、中特征图和大特征图,将三个尺度的特征图输送至多尺度预测框中,多尺度预测框对三个尺度的特征图分别生成预测框,预测框内即为目标检测结果。

[0049] 具体实施方式二:本实施方式对实施方式一作进一步说明,可可变占比图像生成器输入的原始数据为:

[0050] $F_{imgin}(x, y) \mid_{0 < x, y < t}$;

[0051] 将原始数据平移得到: $F_{imgin}(x, y) \mid_{t1 < x, y < t+t1}$;

[0052] 其中, $t1$ 表示填充的黑色边带的宽度,受多尺度损失反馈调节,填充后的图片函数为:

[0053] $F_{imgout}(x, y) \mid_{v=0 < x, y < t+2*t1} = F_{imgin}(x, y) \mid_{\omega=t1 < x, y < t+t1} + (0, 0, 0) \mid_{v-\omega}$;

[0054] 其中, $(0, 0, 0) \mid_a$ 为在 a 区域填充黑色; a 表示以距离原图像 $t1$ 的直线为边界的内侧区域, ω 表示填充前的定义域边界, v 表示填充后的定义域边界,

[0055] 输出的目标数据为:

[0056] $F_{\text{out}}(x, y)_{0 < x, y < t} = F_{\text{imgout}}(x / [(t + 2 * t1) / t], y / [(t + 2 * t1) / t]) \mid_{v=0 < x < t + 2 * t1};$

[0057] (x, y) 表示二维平面上点的位置。

[0058] 本实施方式中, (x, y) 表示二维平面上点的位置, $F(x, y)$ 表示图像的二维离散表示函数, 对应点的函数值是这一位置的亮度值。

[0059] 本实施方式中, 针对两个数据集中小目标比较多的特点, 以及原始存在的数据增强对小目标并没有较好的改善作用, 因此在训练阶段应用目标可变占比数据增强算法改善对小目标的精度。该数据增强的原理是随机的在训练数据图像的四周填充黑色边带, 然后将填充后的图像进行resize使得与要求大小一致, 其结果发生了改变, 首先目标的位置发生了改变, 其次, 目标占图像大小的百分比变小, 即生成了更多的小目标, 目标尺寸缩小到原来的 $\frac{t}{t + 2 * t1}$ 倍, 当数据经过算法之后输入到神经网络, 神经网络学习更多的小目标, 从而

提高神经网络对小目标检测的精度。由缩小比例 $\frac{t}{t + 2 * t1}$ 可知, 当 $t1$ 越小, 则生成的目标越小, 利用损失反馈回来的信息来调节 $t1$, 从而提高算法对不同尺度的目标的检测能力。

[0060] 具体实施方式三: 本实施方式对实施方式二作进一步说明, 可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t + 2 * t1}$ 倍。

[0061] 具体实施方式四: 下面结合图1说明本实施方式, 本实施方式对实施方式三作进一步说明, 该目标检测方法还包括: 损失函数计算模块和多尺度损失判别器;

[0062] 损失函数计算模块分别对三个尺度的特征图计算损失函数, 多尺度损失判别器根据损失函数的大小调节可变占比图像生成器的填充区域, 当小特征图无法获得目标检测结果时, 增大 $t1$, 使得可变占比图像生成器输出的目标数据是将输入数据缩小到原来的 $\frac{t}{t + 2 * t1}$ 倍。

[0063] 本实施方式中, 损失函数(loss)是深度学习分类回归任务的优化目标, 是衡量预测结果和目标差异的指标。

[0064] 具体实施方式五: 本实施方式对实施方式二作进一步说明, 特征提取网络包括五个残差单元;

[0065] 可变占比图像生成器的输出数据 $F_{\text{out}}(x, y)_{0 < x, y < t}$ 经过五个残差单元的五次下采样运算获取小特征图预测所需的特征图 $f1$;

[0066] $f1$ 经过多尺度池化层和上采样, 然后与中浅特征层四次下采样的结果进行拼接, 获得中特征图预测所需的特征图 $f2$;

[0067] $f2$ 经过多尺度池化层和上采样, 然后与浅特征层四次下采样的结果进行拼接, 获得大特征图预测所需的特征图 $f3$ 。

[0068] 具体实施方式六: 下面结合图3说明本实施方式, 本实施方式对实施方式五作进一步说明, 所述多尺度特征生成器产生三个尺度特征图的具体方法包括:

[0069] 将提取的特征经过多尺度特征池化模块, 捕捉不同尺度的特征, 然后将不同尺度的特征经过多尺度特征融合;

[0070] 多尺度特征包括：

[0071] 将大特征图预测所需的特征图f3的特征经过下采样后与中特征图预测所需的特征图f2进行融合,获得中特征图,

[0072] 将中特征图预测所需的特征图f2的特征经过下采样后与小特征图预测所需的特征图f1进行融合,获得小特征图,

[0073] 大特征图预测所需的特征图f3的特征不变,获得大特征图。

[0074] 本实施方式中,将特征通过不同的最大池化后进行拼接,由于把一个feature map从不同的尺度进行特征提取,再聚合的特点,能够捕捉不同尺度下的特征,提高了算法对不同尺度目标的敏感度,进而提高算法的鲁棒性。此外,该模块能够扩大神经网络的感受野,而更大感受野往往能够使得网络能够学习到更好的特征。

[0075] 本实施方式中,原始算法由backbone输出大小为13*13特征,13*13的特征经过上采样后与Res8输出的特征进行拼接进而得到26*26的特征图,同理得到52*52的特征图,最后在三个特征图上进行预测,在13*13的特征图上,具有较大的感受野,更利于检测到较大的目标,而在52*52的特征图上,由于特征图较大,感受野较小,对小目标较敏感。但较小特征图忽略了较大特征图中的浅层特征和包含的其它尺度目标的信息。本文提出了多尺度特征融合,如图3所示,利用将较大特征图的特征经过下采样之后与较小特征图进行融合,使得三个尺度上的特征都能够保留深浅层特征和特征尺度多样性,从而改善特征的代表能力,提高小目标检测的精度和复杂场景的鲁棒性。

[0076] 具体实施方式七:本实施方式对实施方式六作进一步说明,所述小特征图的大小为13*13,中特征图的大小为26*26,大特征图的大小为52*52。

[0077] 具体实施方式八:本实施方式对实施方式六作进一步说明,所述多尺度特征池化模块包括四个不同大小的池化结构,能够捕捉四个不同尺度的特征,四个不同大小池化结构的最大池化结构max_pooli表示i*i最大池化,此处i=1,5,9,13。

[0078] 本实施方式中,采用四个不同大小的池化结构,能够捕捉不同尺度的有效特征,能够使得提取的特征的鲁棒性更好。

[0079] 具体实施方式九:下面结合图2说明本实施方式,本实施方式对实施方式八作进一步说明,所述多尺度特征池化模块的函数表达式为:

[0080] $F_{MSP_out} = \text{concat}(\text{max_pool1}(F_{MSP_in}), \text{max_pool5}(F_{MSP_in}),$

[0081] $\text{max_pool9}(F_{MSP_in}), \text{max_pool13}(F_{MSP_in}))$

[0082] 其中, F_{MSP_in} 表示多尺度特征池化模块的输入特征, F_{MSP_out} 表示多尺度特征池化模块的输出特征,concat表示对特征张量进行拼接。

[0083] 具体实施方式十:本实施方式对实施方式六作进一步说明,多尺度预测框对三个尺度的特征图分别生成预测框的具体方法包括:

[0084] 多尺度预测框在每个尺度的特征图的每个网格上设置三个先验框,获得预测值,每个预测值是(4+1+n)维向量,4表示位置信息,1表示边框置信度,n表示n类目标的概率,n类目标是数据集中包含的目标类别;

[0085] 对预测值与人工标注框计算交并比,保留交并比大于阈值的预测值,如交并比均小于阈值,则保留边框置信度排序为前三位的预测值;

[0086] 对保留的预测值进行非极大值抑制,去除重叠大于阈值的边框,获得预测框。

[0087] 本实施方式中,n类目标是数据集中包含的目标类别,例如coco数据集中有交通工具等80类目标,在使用coco数据集时n的值即为80。

[0088] 虽然在本文中参照了特定的实施方式来描述本发明,但是应该理解的是,这些实施例仅仅是本发明的原理和应用的示例。因此应该理解的是,可以对示例性的实施例进行许多修改,并且可以设计出其他的布置,只要不偏离所附权利要求所限定的本发明的精神和范围。应该理解的是,可以通过不同于原始权利要求所描述的方式来结合不同的从属权利要求和本文中所述的特征。还可以理解的是,结合单独实施例所描述的特征可以使用在其他所述实施例中。

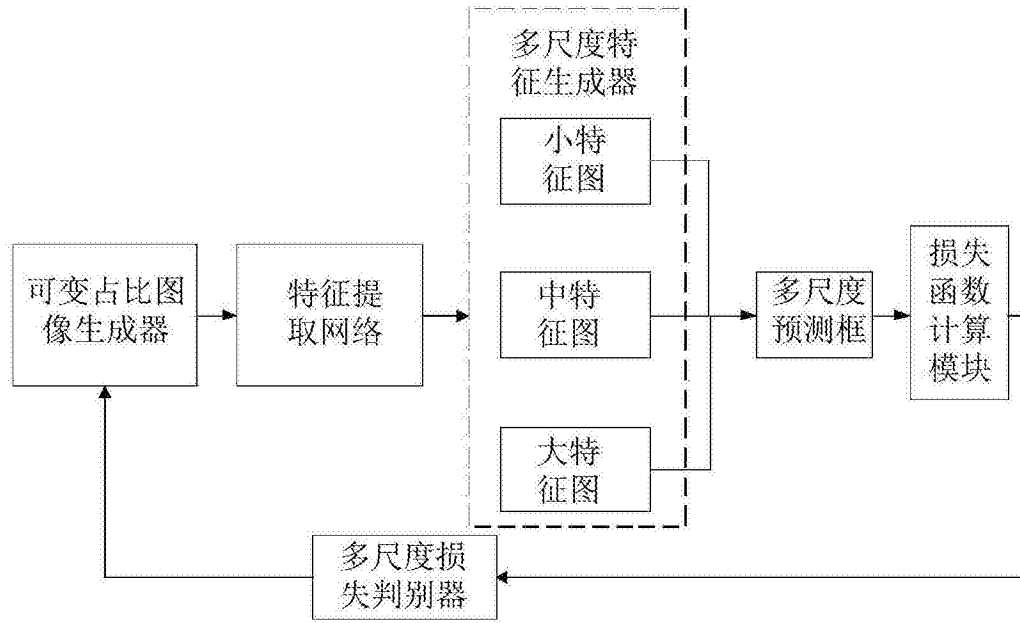


图1

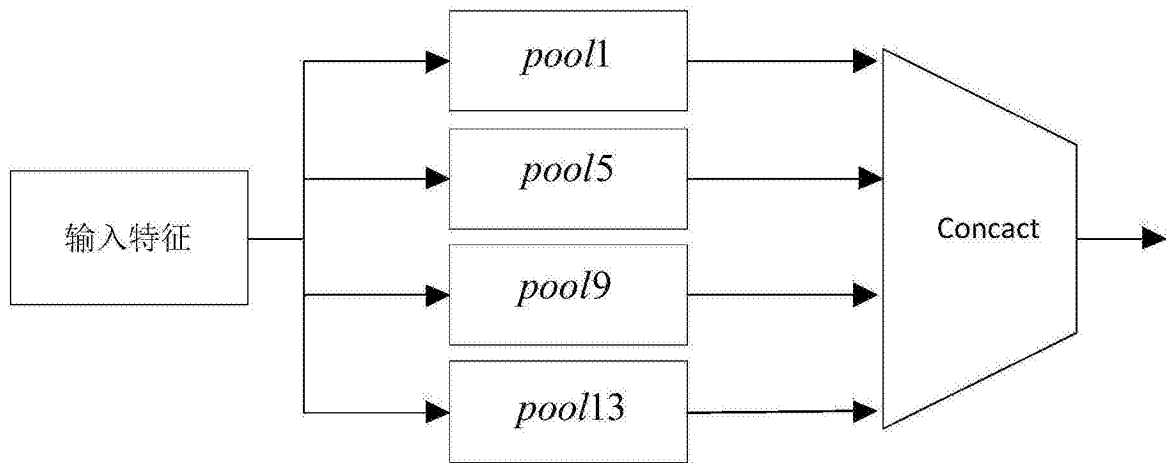


图2

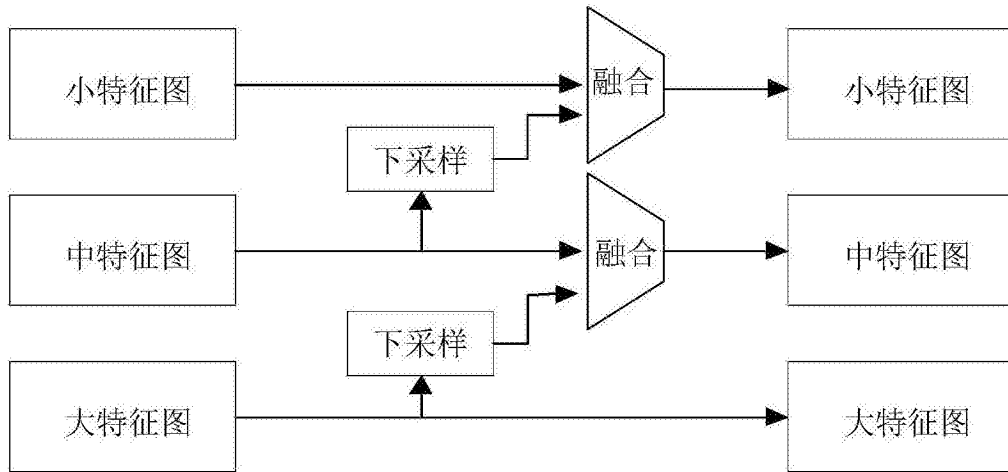


图3