



(12) 发明专利申请

(10) 申请公布号 CN 112785626 A

(43) 申请公布日 2021.05.11

(21) 申请号 202110111717.3

G06N 3/08 (2006.01)

(22) 申请日 2021.01.27

(71) 申请人 安徽大学

地址 230601 安徽省合肥市经开区九龙路
111号

(72) 发明人 涂铮铮 朱庆文 李成龙 汤进
罗斌

(74) 专利代理机构 南京华恒专利代理事务所
(普通合伙) 32335

代理人 宋方园

(51) Int. Cl.

G06T 7/246 (2017.01)

G06T 7/73 (2017.01)

G06K 9/62 (2006.01)

G06N 3/04 (2006.01)

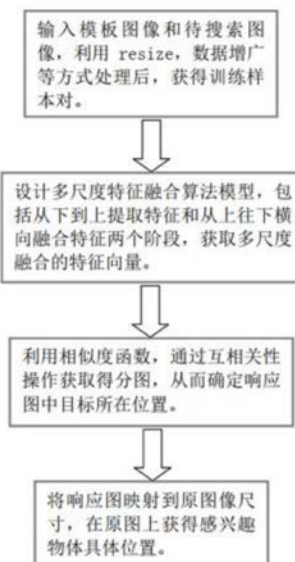
权利要求书2页 说明书5页 附图2页

(54) 发明名称

一种基于多尺度特征融合的孪生网络小目标跟踪方法

(57) 摘要

本发明公开一种基于多尺度特征融合的孪生网络小目标跟踪方法,多尺度融合特征模块以及优化的孪生神经网络全面考虑到深度神经网络结构中低层有利于目标的精确位置,高层可以捕获目标的语义信息的优势,通过不同层次的有效融合,充分利用底层信息避免了深层网络的卷积操作会将小目标的信息抛弃的问题,解决了跟踪过程中的小目标挑战,从而实现了良好的跟踪效果。



1. 一种基于多尺度特征融合的孪生网络小目标跟踪方法,其特征在於:包括以下步骤:

步骤(1)、对模板图像x和待搜索图像y两个图像分别依次进行修改尺寸和数据增广处理,获得对应大小固定的裁剪后的训练样本对,分别输入孪生网络结构中模板分支和搜索分支;

步骤(2)、模板分支和搜索分支共享特征提取器,即使用多尺度特征融合模块来获取多尺度融合特征向量,包括从下到上提取特征和从上往下横向融合特征两个阶段;

从下到上提取特征时构建优化孪生网络结构,该优化孪生网络结构包括5个卷积层,每层的输出依次记为{C1,C2,C3,C4,C5};

从上往下横向融合特征时先对高层特征进行上采样扩大尺寸后与较低一层的特征融合,然后迭代分别生成模板分支与待搜索分支的多尺度融合后的特征图;

步骤(3)、将步骤(2)所得模板特征图和搜索特征图输入相似度函数,进行相关交叉操作获取响应图,响应图中分值较高的位置则被认定为两幅图像目标物体最相似的位置,从而确定待搜索图像y中目标所在位置;

步骤(4)、将响应图扩大到原待搜索图像y的尺寸,然后分析响应图得到最终跟踪结果,将得分最大的位置乘以优化孪生网络结构五层卷积的总步长,即可得到当前目标在待搜索图像上的位置信息。

2. 根据权利要求1所述的基于多尺度特征融合的孪生网络小目标跟踪方法,其特征在於:所述步骤(1)中对模板图像x修改尺寸的具体方法如下:

设第一帧目标框的大小为(x_min,y_min,w,h);然后根据第一帧目标框来计算模板图像x的大小,即以需要追踪的目标为中心裁剪出一个正方形区域,计算公式如下:

$$s(w+2p) \times s(h+2p) = A$$

$$p = \frac{w+h}{4}$$

其中,s是指修改尺寸变换,A设定为127*127大小;通过以上操作将目标框大小扩展,然后修改尺寸变换到127*127大小以获得模板图像x;

对待搜索图像y修改尺寸的具体方法:

先根据上一帧预测的目标框的中心为裁剪中心,然后根据模板图像x裁剪出的正方形区域边长并按比例确定搜索框的边长;最后修改尺寸到255*255大小。

3. 根据权利要求1所述的基于多尺度特征融合的孪生网络小目标跟踪方法,其特征在於:所述步骤(1)中数据增广的方法包括为了增加深度学习训练数据,这里我们利用四种数据增广方式:randomstretch随机拉伸,randomcrop随机裁剪,normalize标准化和totensor转化为张量;

最后修改尺寸到需要输入到网络结构中的大小。

4. 根据权利要求1所述的基于多尺度特征融合的孪生网络小目标跟踪方法,其特征在於:所述步骤(2)中构建优化孪生网络结构来从下到上提取特征,优化孪生网络结构设置如下:

①、第一层为卷积层,使用11*11*96卷积核,步长为2,对图像进行卷积操作.然后使用3*3的最大池化操作和批标准化操作,输出C1;

②、第二层为卷积层,使用5*5*256,步长为1的卷积核使用两组GPU分别进行卷积操作,

然后使用3*3的最大池化操作和批标准化操作来提取特征信息,输出C2;

③、第三层为卷积层,使用3*3*192的卷积核分组进行卷积操作并继续批标准化操作,输出C3;

④、第四层为卷积层,使用3*3*192的卷积核分组进行操作并继续批标准化操作,输出C4;

⑤、第五层为卷积层,仅使用3*3*128的卷积操作,最后的输出256维的高层语义特征C5。

5. 根据权利要求1所述的基于多尺度特征融合的双生网络小目标跟踪方法,其特征在于:所述步骤(2)中从上往下横向融合特征的具体方法为:

(A)、采用内插值法先在第五层的特征图像素基础上采用2倍上采样在像素之间插入新的元素,将其大小变为第四层的特征尺寸,从而扩大高层特征大小便于下一步融合;然后依次扩大第四层、第三层以及第二层的特征大小;

(B)、在C5层使用一个1*1的卷积操作得到低分辨率的特征P5,然后利用一个1*1的卷积核改变自下而上过程中生成的第四层特征图C4的通道数,将其通道统一固定为256-d,接下来将第四层处理后的结果与第五层进行的采样后结果进行相加融合,并使用一个3*3的卷积核处理融合后的结果,将最后得到的结果记作P4;

迭代上述(B)过程最终生成特征图,分别获得模板分支与待搜索分支的多尺度融合后的特征图。

6. 根据权利要求1所述的基于多尺度特征融合的双生网络小目标跟踪方法,其特征在于:所述步骤(3)将模板分支与待搜索分支对应的多尺度融合后的特征图,利用互相关性操作获取响应图;

互相关操作的具体过程为:利用模板分支与待搜索分支对应多尺度融合后的特征,这两个特征的尺寸分别为22*22*256和6*6*256,接着将6*6*256作为一个卷积核在22*22*256的特征上进行卷积操作,获得一个17*17的响应图;

在训练过程中,获得17*17响应图后接着确定正负样本:在搜索图像上只要距离目标的值小于R,则算为正样本,反之则认为负样本;

最后,采用二分类交叉熵逻辑损失函数,利用随机梯度下降法进行迭代训练,训练整个深度网络;

上述相似度函数公式如下: $f(z, x) = \varphi(z) * \varphi(x) + b_1$;

其中, $\varphi(z)$ 为卷积核,在 $\varphi(x)$ 上进行卷积, b_1 表示得分图上每个位置的取值。

一种基于多尺度特征融合的孪生网络小目标跟踪方法

技术领域

[0001] 本发明涉及视觉识别技术,具体涉及一种基于多尺度特征融合的孪生网络小目标跟踪方法。

背景技术

[0002] 运动物体跟踪是指在给定一段视频序列的第一帧的感兴趣目标的位置信息后,跟踪器能在后续的序列中继续精确地、实时地跟踪目标,返回位置信息。近几年,目标跟踪的理论方法发展非常迅速,它是计算机视觉领域的一个重要的研究方向,并且已经被成功地应用于视频监控、无人驾驶、语义分割等多个领域。深度学习方法的出现极大的促进了跟踪问题的发展,但是小目标跟踪问题仍是一个非常大的挑战,特别是在复杂的场境中如何实时地、精确地追踪小目标是研究的重点问题。

[0003] 目前,小目标跟踪的挑战性主要来源于两方面:小目标物体随着神经网络深度的增加其特征非常难以获取,因此获取特征表示困难。另一方面,在跟踪过程中,由于镜头抖动,与正常尺寸的目标相比,小目标往往会发生突然的大幅度的漂移。目前的研究仅仅关注于通用数据集上的正常尺寸目标物体的跟踪结果,但是却忽略了小目标跟踪问题。

[0004] 现有的小目标跟踪算法都是基于传统的机器学习算法,无论是精度的提升或跟踪的实时性上都存在很大的局限性,而深度神经网络由于其较深的网络层数能够提取高层的语义信息从而更好的表达特征,但是对于小目标的物体来说,随着网络层数的加深,不断的卷积操作会导致网络逐渐丢失小目标的位置信息。

[0005] 因此,利用孪生网络的深度神经网络结构,可从多尺度特征融合角度出发,通过融合不同网络层互补的特征信息,实现复杂场景和环境下实时性、鲁棒性的小目标物体跟踪,但现有孪生网络的应用存在以下问题:如何有效融合不同网络层的多尺度的特征、现有深度神经网络目标位置模糊以及语义信息较少等,最终导致难以获取小目标特征。

发明内容

[0006] 发明目的:本发明的目的在于解决现有技术中存在的不足,提供一种基于多尺度特征融合的孪生网络小目标跟踪方法,本发明中独创有类似特征金字塔的基于全卷积孪生网络的多尺度特征融合模块来增加网络对尺度变化的鲁棒性,使得端到端的训练从而实现精准地、实时地追踪小尺度目标。

[0007] 技术方案:本发明一种基于多尺度特征融合的孪生网络小目标跟踪方法,包括以下步骤:

[0008] 步骤(1)、对模板图像 x 和待搜索图像 y 两个图像分别依次进行修改尺寸和数据增广预处理,获得对应大小固定的裁剪后的训练样本对,分别输入孪生网络结构中模板分支和搜索分支;

[0009] 步骤(2)、模板分支和搜索分支共享特征提取器,即使用多尺度特征融合模块来获取多尺度融合特征向量,包括从下到上提取特征和从上往下横向融合特征两个阶段;

[0010] 从下到上提取特征时构建优化孪生网络结构,该优化孪生网络结构包括5个卷积层,每层的输出依次记为{C1,C2,C3,C4,C5};

[0011] 从上往下横向融合特征时先对高层特征进行上采样扩大尺寸后与较低一层的特征融合,然后迭代分别生成模板分支与待搜索分支的多尺度融合后的特征图;

[0012] 步骤(3)、将步骤(2)所得模板特征图和搜索特征图输入相似度函数,进行相关交叉操作获取响应图,响应图中分值较高的位置则被认定为两幅图像目标物体最相似的位置,从而确定目标所在位置,即是指待搜索图像(即需要跟踪的帧)中目标位置;

[0013] 步骤(4)、将响应图扩大到原待搜索图像 y 尺寸(例如225*225),然后分析响应图得到最终跟踪结果,将得分最大的位置乘以优化孪生网络结构五层卷积的总步长,即可得到当前目标在待搜索图像上的位置信息。

[0014] 进一步的,所述步骤(1)中对模板图像 x 修改尺寸的具体方法如下:

[0015] 目标跟踪过程中第一帧目标框的大小是已知的,设第一帧目标框的大小为(x_{\min}, y_{\min}, w, h);然后根据第一帧目标框来计算模板图像 x 的大小,即以需要追踪的目标为中心裁剪出一个正方形区域,计算公式如下:

$$[0016] \quad s(w+2p) \times s(h+2p) = A$$

$$[0017] \quad p = \frac{w+h}{4}$$

[0018] 其中, (x_{\min}, y_{\min}) 是指目标框的左下角的坐标值, w 和 h 表示框的宽和高, s 是指修改尺寸, A 设定为127*127大小;通过以上操作将目标框大小扩展,然后修改尺寸到127*127大小以获得模板图像 x 。

[0019] 本发明将一段视频帧中的第一帧叫做模板帧(即模板图像 x),后续帧都是待搜索目标位置的(即待搜索图像 y),位置均用左下角和宽高四个坐标表示。

[0020] 对待搜索图像 y 进行修改尺寸的具体方法:

[0021] 先根据上一帧预测的目标框的中心为裁剪中心,然后根据模板图像 x 裁剪出的正方形区域边长并按比例确定搜索框的边长;最后修改尺寸到255*255大小。

[0022] 进一步的,所述步骤(2)中构建优化孪生网络结构来从下到上提取特征,优化孪生网络结构设置如下:

[0023] ①、第一层为卷积层,使用11*11*96卷积核,步长为2,对图像进行卷积操作.然后使用3*3的最大池化操作和批标准化操作,输出C1;

[0024] ②、第二层为卷积层,使用5*5*256,步长为1的卷积核使用两组GPU分别进行卷积操作,然后使用3*3的最大池化操作和批标准化操作来提取特征信息,输出C2;

[0025] ③、第三层为卷积层,使用3*3*192的卷积核分组进行卷积操作并继续批标准化操作,输出C3;

[0026] ④、第四层为卷积层,使用3*3*192的卷积核分组进行操作并继续批标准化操作,输出C4;

[0027] ⑤、第五层为卷积层,仅使用3*3*128的卷积操作,最后的输出256维的高层语义特征C5。

[0028] 进一步的,所述步骤(2)中从上往下横向融合特征的具体方法为:

[0029] (A)、采用内插值法先在第五层的特征图像素基础上采用2倍上采样(最近邻上采

样法)在像素之间插入新的元素,将其大小变为第四层的特征尺寸,从而扩大高层特征大小便于下一步融合;然后依次扩大第四层、第三层以及第二层的特征大小;

[0030] (B)、在C5层使用一个1*1的卷积操作得到低分辨率的特征P5,然后利用一个1*1的卷积核改变自下而上过程中生成的第四层特征图C4的通道数,将其通道统一固定为256-d,便于后续的特征融合接下来将第四层处理后的结果与第五层进行的采样后结果进行相加融合,并使用一个3*3的卷积核处理融合后的结果以解决上采样过程中可能产生的混叠效应,将最后得到的结果记作P4;

[0031] 迭代上述(B)过程最终生成更加精确的特征图,分别获得模板分支与待搜索分支的多尺度融合后的特征图。

[0032] 进一步的,所述步骤(3)将模板分支与待搜索分支对应的多尺度融合后的特征图,利用互相关操作获取响应图。互相关操作,具体操作来说,利用模板分支与待搜索分支对应多尺度融合后的特征,这两个特征的尺寸分别为22*22*256和6*6*256,接着将6*6*256作为一个卷积核在22*22*256的特征上进行卷积操作,获得一个17*17的响应图,此17*17的响应图上跟踪的目标位置处分值较高;

[0033] 在训练过程中,获得17*17响应图后接着确定正负样本:在搜索图像上只要距离目标的值小于R,则算为正样本,反之则认为负样本;

[0034] 最后,采用二分类交叉熵逻辑损失函数,利用随机梯度下降法,训练迭代次数设为50,最小批次设置为8,学习率从 10^{-2} 衰减为 10^{-8} 训练整个深度网络;

[0035] 上述相似度函数公式如下: $f(z, x) = \varphi(z) * \varphi(x) + b_1$;

[0036] 其中, $\varphi(z)$ 为卷积核,在 $\varphi(x)$ 上进行卷积, b_1 表示得分图上每个位置的取值。

[0037] 有益效果:本发明中设有多尺度融合特征模块,全面考虑到深度神经网络结构中低层有利于目标的精确位置,高层可以捕获目标的语义信息的优势,通过不同层次的有效融合,充分利用底层信息避免深层网络的卷积操作会将小目标的信息抛弃的问题。此外,本发明优化现有孪生网络结构,能够精确实时地追踪小目标物体的视觉目标跟踪方法。

[0038] 综上所述,本发明能全面有效融合不同网络层特征的结构,解决了跟踪过程中的小目标挑战,从而实现了良好的跟踪效果。

附图说明

[0039] 图1为本发明的整体流程示意图;

[0040] 图2为本发明一实施例中待搜索分支的多尺度特征融合模块结构示意图;

[0041] 图3为本发明中一实施例对比示意图;

[0042] 其中,图3(a)为采用本发明所得可视化特征图,图3(b)为采用现有孪生网络所得可视化特征图。

具体实施方式

[0043] 下面对本发明技术方案进行详细说明,但是本发明的保护范围不局限于所述实施例。

[0044] 在目标跟踪的实际应用中,存在中高度的海拔下对相机拍摄的目标进行跟踪的情

况,在远距离的场景中如何持续精确地追踪目标一直是跟踪领域研究的难点问题。

[0045] 本发明基于优化后的孪生网络,通过自上而下的多尺度融合方法进行特征融合,解决了现有技术中小物体跟踪困难的问题,如图1所示,本发明一种基于多尺度特征融合的孪生网络小目标跟踪方法,包括以下步骤:

[0046] 步骤(1)、对模板图像 x 和待搜索图像 y 两个图像分别依次进行修改尺寸数据增广的预处理,获得对应大小固定的裁剪后的训练样本对,分别输入孪生网络结构中模板分支和搜索分支;

[0047] 目标跟踪过程中,设第一帧目标框的大小为 (x_min, y_min, w, h) ;然后根据第一帧目标框来计算模板图像 x 的大小,即以需要追踪的目标为中心裁剪出一个正方形区域,计算公式如下:

$$[0048] \quad s(w+2p) \times s(h+2p) = A$$

$$[0049] \quad p = \frac{w+h}{4}$$

[0050] 其中, s 是指修改尺寸, A 设定为 $127*127$ 大小;通过以上操作将目标框大小扩展,然后修改尺寸到 $127*127$ 大小以获得模板图像 x ;

[0051] 训练中,对待搜索图像 y 修改尺寸的具体方法:

[0052] 先根据上一帧预测的目标框的中心为裁剪中心,然后根据模板图像 x 裁剪出的正方形区域边长并按比例确定搜索框的边长;最后修改尺寸到 $255*255$ 大小;

[0053] 步骤(2)、模板分支和搜索分支共享特征提取器,即使用多尺度特征融合模块来获取多尺度融合特征向量,包括从下到上提取特征和从上往下横向融合特征两个阶段;

[0054] 如图2所示,构建优化孪生网络结构来从下到上提取特征,优化孪生网络结构设置如下:

[0055] ①、第一层为卷积层,使用 $11*11*96$ 卷积核,步长为2,对图像进行卷积操作.然后使用 $3*3$ 的最大池化操作和批标准化操作,输出 $C1$;

[0056] ②、第二层为卷积层,使用 $5*5*256$,步长为1的卷积核使用两组GPU分别进行卷积操作,然后使用 $3*3$ 的最大池化操作和批标准化操作来提取特征信息,输出 $C2$;

[0057] ③、第三层为卷积层,使用 $3*3*192$ 的卷积核分组进行卷积操作并继续批标准化操作,输出 $C3$;

[0058] ④、第四层为卷积层,使用 $3*3*192$ 的卷积核分组进行操作并继续批标准化操作,输出 $C4$;

[0059] ⑤、第五层为卷积层,仅使用 $3*3*128$ 的卷积操作,最后的输出256维的高层语义特征 $C5$ 。

[0060] 从上往下横向融合特征的具体方法为:

[0061] (A)、采用内插值法先在第五层的特征图像素基础上采用2倍上采样(最近邻上采样法)在像素之间插入新的元素,将其大小变为第四层的特征尺寸,从而扩大高层特征大小便于下一步融合;然后依次扩大第四层、第三层以及第二层的特征大小;

[0062] (B)、在 $C5$ 层使用一个 $1*1$ 的卷积操作得到低分辨率的特征 $P5$,然后利用一个 $1*1$ 的卷积核改变自下而上过程中生成的第四层特征图 $C4$ 的通道数,将其通道统一固定为256-d,便于后续的特征融合接下来将第四层处理后的结果与第五层进行的采样后结果进行相

加融合,并使用一个3*3的卷积核处理融合后的结果以解决上采样过程中可能产生的混叠效应,将最后得到的结果记作P4;

[0063] 迭代上述(B)过程最终生成更加精确的特征图,分别获得模板分支与待搜索分支的多尺度融合后的特征图;

[0064] 步骤(3)、所述步骤(3)将模板分支与待搜索分支对应的多尺度融合后的特征图,利用互相关操作获取响应图。互相关操作,具体操作来说,利用模板分支与待搜索分支对应多尺度融合后的特征,这两个特征的尺寸分别为22*22*256和6*6*256,接着将6*6*256作为一个卷积核在22*22*256的特征上进行卷积操作,获得一个17*17的响应图,此17*17的响应图上跟踪的目标位置处分值较高;

[0065] 在训练过程中,获得响应图后需确定正负样本:在搜索图像上只要距离目标的值小于R,则算为正样本,反之则认为负样本;

[0066] 最后,采用二分类交叉熵逻辑损失函数,利用随机梯度下降法,训练迭代次数设为50,最小批次设置为8,学习率从 10^{-2} 衰减为 10^{-8} 训练整个深度网络;

[0067] 上述相似度函数公式如下: $f(z, x) = \varphi(z) * \varphi(x) + b_1$;

[0068] 其中, $\varphi(z)$ 为卷积核,在 $\varphi(x)$ 上进行卷积, b_1 表示得分图上每个位置的取值;

[0069] 步骤(4)、将响应图扩大到原图像尺寸,然后分析响应图得到最终跟踪结果,将得分最大的位置乘以优化孪生网络结构五层卷积的总步长,即可得到当前目标在待搜索图像上的位置信息。

[0070] 如图3所示,采用本发明方法所得目标定位精准,效果更清晰。

[0071] 通过上述实施例可以看出,本发明将目标跟踪看作相似性度量问题的学习。对于模板图像x和待搜索图像y,我们将其输入到孪生网络结构中进行相同的变换,并设计多尺度特征融合模块分别获取对应特征向量,最后将模板特征图当成卷积核在搜索特征上进行互相关性操作,生成响应图从而比较两者之间的相似度,相似度较高的位置则返回一个高分值,即目标位置,否则返回一个低分值。

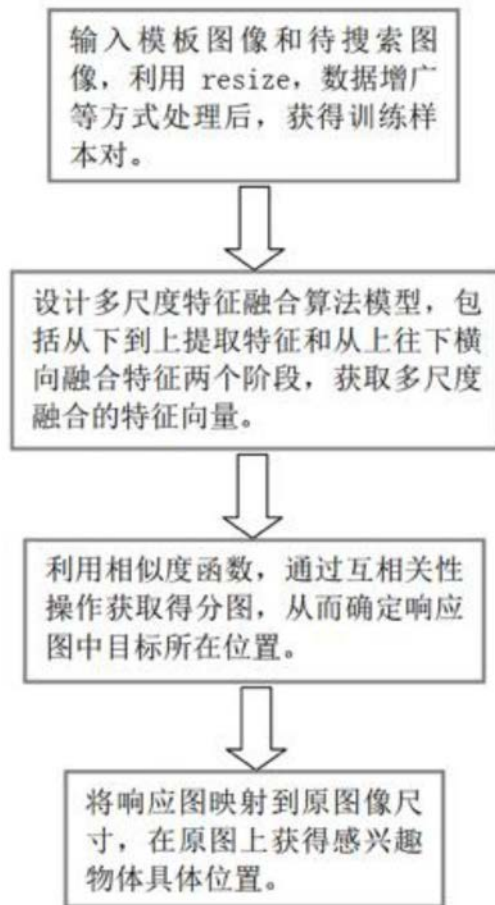


图1

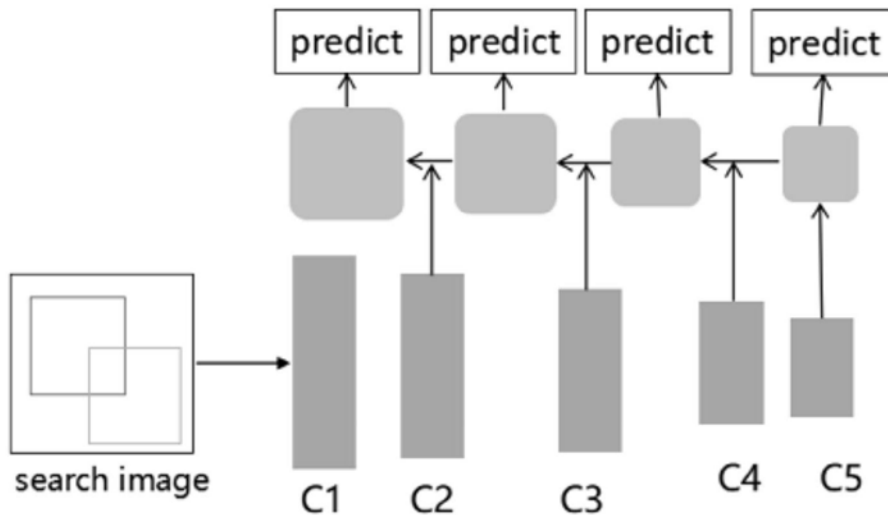


图2



3 (a)



3 (b)

图3