(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 110929848 A
(43)申请公布日 2020.03.27

(21)申请号 201911128508.9

(22)申请日 2019.11.18

(71)申请人 安徽大学
地址 230000 安徽省合肥市蜀山区肥西路3号

(72)发明人 李成龙 刘磊 鹿安东

(74)专利代理机构 合肥市浩智运专利代理事务所(普通合伙) 34124
代理人 张景云

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

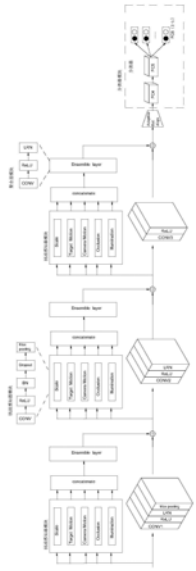
权利要求书3页 说明书9页 附图3页

(54)发明名称

基于多挑战感知学习模型的训练、跟踪方法

(57)摘要

本发明涉及基于多挑战感知学习模型的训练及实时跟踪方法,依次包括模型训练过程、通过预先训练的模型进行的跟踪过程两个部分,其中,S11、构建网络模型S12、使用标定好目标的VOT数据集来训练整个所述的网络模型;S21、输入当前跟踪的视频帧,在前一帧预测的目标位置周围用高斯采样获取当前帧的候选样本;S22、获取候选样本的的特征图;S23、将所述特征图输入到分类器模块中,预测目标位置;S24、判断当前帧是否跟踪成功;本发明能够有效的增加特征表达的丰富性,提高了跟踪的鲁棒性,并达到了实时的跟踪性能。



CN 110929848 A

1.基于多挑战感知学习模型的训练方法,其特征在于,包括以下步骤:

S11、构建网络模型;

所述网络模型由依次串联的用于获取候选样本特征图的多级挑战模块、Adaptive RoI Align层、分类器模块组成;

S12、使用标定好目标的VOT数据集来训练所述的网络模型。

2.根据权利要求1所述的基于多挑战感知学习模型的模型的训练方法,其特征在于,所述步骤S11中;

所述多级挑战模块包括第一级挑战模块、第二级挑战模块、第三级挑战模块;

所述第一级挑战模块包括第一卷积层模块、第一多挑战感知器模块、第一concatenate函数层、第一聚合层模块,将候选样本分别输入至第一卷积层模块、第一多挑战感知器模块中,第一卷积层模块用来提取通用的目标特征,第一多挑战感知器模块提取不同挑战属性下的目标特征表示,然后第一多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接,并输送至第一聚合层模块,第一聚合层模块将得到的多挑战特征进行聚合处理,第一聚合层模块处理的结果与第一卷积层模块提取的目标特征进行相加融合,输送至所述第二级挑战模块处;

所述第二级挑战模块包括第二卷积层模块、第二多挑战感知器模块、第二concatenate函数层、第二聚合层模块,所述第二卷积层模块、第二多挑战感知器模块接收到第一级挑战模块输出的相加融合的结果;第二多挑战感知器模块提取不同挑战属性下的目标特征表示,然后第二多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接,并输送至第二聚合层模块,第二聚合层模块将得到的多挑战特征进行聚合处理,第二聚合层模块处理的结果与第二卷积层模块提取的目标特征进行相加融合;输送至第三级挑战模块处;

所述第三级挑战模块包括第三卷积层模块、第三多挑战感知器模块、第三concatenate函数层、第三聚合层模块,所述第三卷积层模块、第三多挑战感知器模块接收到所述第二级挑战模块输出的相加融合的结果;第三多挑战感知器模块提取不同挑战属性下的目标特征表示,然后第三多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接,并输送至第三聚合层模块,第三聚合层模块将得到的多挑战特征进行聚合处理,第三聚合层模块处理的结果与第三卷积层模块提取的目标特征进行相加融合;

输送到达Adaptive RoI Align层,Adaptive RoI Align层加快跟踪过程中候选区域的特征提取过程,根据不同候选样本提取对应位置的深度特征得到最终的特征图,再输送至分类器模块,分类器模块进行计算得到每个候选样本的得分。

3.根据权利要求2所述的基于多挑战感知学习模型的训练方法,其特征在于,

所述第一卷积层模块、第二卷积层模块、第三卷积层模块作为主干网络模块,通过使用ImageNet数据集进行预训练分类网络VGG-M,并将这个网络的原有参数作为主干网络模块的初始化参数;

所述多挑战感知器模块由依次串联的卷积层、ReLU激活函数层、批归一化层、随机失活层、最大池化层组成;

所述聚合层模块由依次串联的卷积层、ReLU激活函数层、局部响应归一化层组成;

所述分类器模块是由依次串联的两个全连接层和一个带由softmax层的全连接层所组

成；

第一卷积层模块由依次串联的第一卷积层、ReLU激活函数层、局部响应归一化层、最大池化层组成；

所述第二卷积层模块由依次串联的第二卷积层、ReLU激活函数层、局部响应归一化层组成；

第三卷积层模块由依次串联的第三卷积层、ReLU激活函数层组成。

4.根据权利要求3所述的基于多挑战感知学习模型的训练方法，其特征在于，所述步骤S12包括；

S1201、在每一帧中根据给定的真值框选取S_+＝50(IOU≥0.7)和S_-＝200(IOU≤0.5)的样本数；其中，S_+表示正样本，S_-表示负样本，IOU表示采集样本与真值框之间的交并比；

S1202、通过采集的正负样本，使用随机梯度下降法进行迭代训练，每次迭代训练根据以下方法处理：VOT数据集中的视频序列的个数为K，K为正整数，并为每个视频序列构建一个新的随机初始化的FC6全连接层；

S1203、训练分为两个阶段，第一阶段提取VOT数据集中每个视频序列不同挑战帧的集合，用不同挑战帧的集合训练多挑战感知器模块；

第二阶段使用整个VOT数据集训练聚合层模块，得到最终的用来进行视觉目标跟踪的训练模型。

5.基于权利要求1-4任一所述的多挑战感知学习模型的训练方法的实时视觉跟踪方法，其特征在于，包括以下步骤：

S21、输入当前跟踪的视频帧，在前一帧预测的目标位置周围用高斯采样获取当前帧的候选样本；

S22、获取候选样本的的特征图；

S23、将所述特征图输入到分类器模块中，预测目标位置；

S24、判断当前帧是否跟踪成功，正样本的得分大于0时，跟踪成功，执行步骤1)；正样本的得分小于0时，即跟踪失败，执行步骤2)。

6.根据权利要求5所述的基于多挑战感知学习模型的实时视觉跟踪方法，其特征在于，所述步骤S21还包括；

由待跟踪视频序列的提供的第一帧图像作为前一帧；由前一帧和框定目标位置区域的真值框，按照高斯分布随机产生样本，使用该样本初始化跟踪模型；

初始化完成后；以前一帧目标位置为均值，以$(0.09r^2,0.09r^2,0.25)$为协方差，通过高斯分布进行采样产生候选样本，其中：r为前一帧目标框的宽和高的平均值。

7.根据权利要求5所述的基于多挑战感知学习模型的实时视觉跟踪方法，其特征在于，所述步骤S22包括；

将候选样本分别输入至多级挑战模块中，直至到达Adaptive RoI Align层，Adaptive RoI Align层加快跟踪过程中候选区域的特征提取过程，根据不同候选样本提取对应位置的深度特征得到最终的特征图。

8.根据权利要求6所述的基于多挑战感知学习模型的实时视觉跟踪方法，其特征在于，所述步骤S23包括；最终的特征图输入分类器模块中，通过分类器模块获得每个候选样本被

判定为正样本和负样本的得分，分别设为$f^+(x^i)$和$f^-(x^i)$，利用公式$x^* = \arg\max_{x^i} f^+(x^i)$确定当前帧的目标位置，其中$x^i$表示采样的第i个样本，$f^+(x^i)$表示获取的正样本得分，$f^-(x^i)$表示获取的负样本得分；$x^*$为预测的目标位置；

得到每个候选样本的得分，最高正样本得分的样本位置作为当前帧预测的视觉跟踪结果。

9.根据权利要求7所述的基于多挑战感知学习模型的实时视觉跟踪方法，其特征在于，

所述步骤1)为：跟踪成功时，在当前帧的预测位置周围采集正样本和负样本，将这些样本以Adaptive RoI Align后的特征保存至总的正负样本数据集中；

所述步骤2)为：跟踪失败时，进行短期更新，短期更新包括：从总正负样本数据集中抽出最近20帧跟踪成功收集的正负样本进行迭代训练。

10.根据权利要求8所述的基于多挑战感知学习模型的实时视觉跟踪装置，其特征在于，所述步骤S24还包括长期更新，其更新规则为固定每隔若干帧执行更新；在对模型进行更新后，判断当前帧是否为最后一帧，若是最后一帧，目标识别及跟踪结束，否则跟踪继续。

基于多挑战感知学习模型的训练、跟踪方法

技术领域

[0001] 本发明涉及计算机视觉领域，尤其涉及基于多挑战感知学习模型的训练、跟踪方法。

背景技术

[0002] 视觉跟踪是计算机视觉领域中一个基础的研究问题，其目的是在给定视频序列第一帧中跟踪目标初始状态(如大小和位置)的情况下，估计后续视频帧中目标的状态。目前，视觉跟踪技术已广泛应用于智能视频监控、无人驾驶、增强现实等领域，对社会安全和文化娱乐等领域的发展有着重要的研究意义。

[0003] 随着计算机硬件性能的不断提高和大规模视觉数据集(如ImageNet，大规模图像分类数据集)的引入，基于深度学习特别是深层卷积神经网络的方法在多个计算机视觉任务上(如图片分类、目标检测)都取得了显著的成功。目前基于深度学习检测方法的视觉跟踪模型，实质上是学习跟踪目标的深度特征表示，再送入一个二分类器中对目标和背景进行分类。但是这种方法在处理视觉跟踪任务时仍具有一定局限性，其中一个关键的原因是深层卷积神经网络算法的性能依赖于大规模标注的训练数据集的离线学习。然而，由于视觉跟踪任务的目标是任意的，很难得到足够的训练数据来学习有效的基于目标实例的深度特征表示。

[0004] 为了解决训练数据不充分的问题，现有的方法在最后一个卷积层后添加了一个Inception-like模块，并使用该模块和VOT(Visual Object Tracking，视觉目标跟踪)数据集中标注的挑战属性来学习基于挑战感知的具有高级语义信息的深度特征表示；如申请号为“CN201710863151.3”的专利，利用了训练模型Inception模块进行处理。

[0005] 然而，我们观察到，一些挑战(如光照变化)在浅层有着很好的特征表示，而一些挑战(如尺度变化)的特征在中层可以表现的很好。因此，现有方法采用的使用Inception-like结构提取不同属性的具有高级语义信息的深度特征表示，不能很好的提取多层次的挑战信息，从而使得挑战属性的特征表达不够丰富。

发明内容

[0006] 本发明所要解决的技术问题在于提供基于多挑战感知学习模型的训练、跟踪方法，以解决多层次的挑战信息提取不佳的问题。

[0007] 本发明通过以下技术手段实现解决上述技术问题的：

[0008] 基于多挑战感知学习模型的训练方法，包括以下步骤；

[0009] S11、构建网络模型；

[0010] 所述网络模型包括依次串联的第一级挑战模块、所述第二级挑战模块、第三级挑战模块、Adaptive RoI Align(自适应感兴趣区域对准操作)层、分类器模块；其中，

[0011] 所述第一级挑战模块包括第一卷积层模块、第一多挑战感知器模块、第一concatenate函数层、第一聚合层模块，将候选样本分别输入至第一卷积层模块、第一多挑

战感知器模块中，第一卷积层模块用来提取通用的目标特征，第一多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第一多挑战感知器模块的输出结果通过第一concatenate(拼接)函数层进行通道维度的拼接，并输送至第一聚合层模块，第一聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第一聚合层模块处理的结果与第一卷积层模块提取的目标特征进行相加融合，输送至所述第二级挑战模块处；

[0012] 所述所述第二级挑战模块包括第二卷积层模块、第二多挑战感知器模块、第二concatenate函数层、第二聚合层模块，所述第二卷积层模块、第二多挑战感知器模块接收到第一级挑战模块输出的相加融合的结果；第二多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第二多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接，并输送至第二聚合层模块，第二聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第二聚合层模块处理的结果与第二卷积层模块提取的目标特征进行相加融合；输送至第三级挑战模块处；

[0013] 所述第三级挑战模块包括第三卷积层模块、第三多挑战感知器模块、第三concatenate函数层、第三聚合层模块，所述第三卷积层模块、第三多挑战感知器模块接收到所述第二级挑战模块输出的相加融合的结果；第三多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第三多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接，并输送至第三聚合层模块，第三聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第三聚合层模块处理的结果与第三卷积层模块提取的目标特征进行相加融合；输送到达Adaptive RoI Align层，Adaptive RoI Align层加快跟踪过程中候选区域的特征提取过程，根据不同候选样本提取对应位置的深度特征得到最终的特征图，再输送至分类器模块，分类器模块进行计算得到每个候选样本的得分，最后取最高正样本得分的样本位置作为当前帧预测的视觉跟踪结果；

[0014] S12、使用标定好目标的VOT数据集来训练整个所述的网络模型；

[0015] 通过引入多层次与主干网络模块并行的多挑战感知器模块学习不同层次的挑战特征，引入聚合层模块来解决跟踪过程中挑战不可知的问题，引入Adaptive RoI Align层来加快跟踪过程中候选区域的特征提取过程；有效的增加了特征表达的丰富性，提高了跟踪的鲁棒性，并达到了实时的跟踪性能。

[0016] 作为本发明进一步的方案：并截取前三层卷积层作为主干网络，所述第一卷积层模块、第二卷积层模块、第三卷积层模块作为主干网络模块，通过使用ImageNet数据集进行预训练分类网络VGG-M，并将这个网络的原有参数作为主干网络模块的初始化参数；

[0017] 所述多挑战感知器模块由依次串联的卷积层、ReLU激活函数层、批归一化层、随机失活层、最大池化层组成；

[0018] 所述聚合层模块由依次串联的卷积层、ReLU激活函数层、局部响应归一化层组成；

[0019] 所述分类器模块是由依次串联的两个全连接层和一个带由softmax层的全连接层所组成。

[0020] 作为本发明进一步的方案：第一卷积层模块由依次串联的第一卷积层、ReLU激活函数层、局部响应归一化层、最大池化层组成；

[0021] 所述第二卷积层模块由依次串联的第二卷积层、ReLU激活函数层、局部响应归一

化层组成；

[0022] 第三卷积层模块由依次串联的第三卷积层、ReLU激活函数层组成；

[0023] 所述第一卷积层、第二卷积层、第三卷积层的卷积核大小分别为7*7、5*5、3*3，第一卷积层、第二卷积层操作步长为2，第三卷积层是操作步长为1、空洞率为3的空洞卷积。

[0024] 作为本发明进一步的方案：所述步骤S12包括；

[0025] S1201、在每一帧中根据给定的真值框选取$S_+=50$(IOU≥0.7)和$S_-=200$(IOU≤0.5)的样本数；其中，S_+表示正样本，S_-表示负样本，IOU表示采集样本与真值框之间的交并比；

[0026] S1202、通过采集的正负样本，使用随机梯度下降法进行1000次迭代训练，每次迭代训练根据以下方法处理：设K表示VOT数据集中的视频序列的个数，K为正整数，为每个视频序列构建一个新的随机初始化的FC6全连接层；

[0027] S1203、训练分为两个阶段，第一阶段提取VOT数据集中每个视频序列不同挑战帧的集合，用这些数据训练与主干网络模块并行的多挑战感知器模块；

[0028] 第二阶段使用整个VOT数据集训练聚合层模块，得到最终的训练模型，用来进行视觉目标跟踪。

[0029] 基于多挑战感知学习模型的实时视觉跟踪方法，包括以下步骤：

[0030] S21、输入当前跟踪的视频帧，在前一帧预测的目标位置周围用高斯采样获取当前帧的候选样本；

[0031] S22、获取候选样本的的特征图；

[0032] S23、将所述特征图输入到分类器模块中，预测目标位置；

[0033] S24、判断当前帧是否跟踪成功，正样本的得分大于0时，跟踪成功，执行步骤1)；正样本的得分小于0时，即跟踪失败，执行步骤2)。

[0034] 为本发明进一步的方案：所述步骤S21还包括；

[0035] 由待跟踪视频序列的提供的第一帧图像作为前一帧；由前一帧和框定目标位置区域的真值框，按照高斯分布随机产生样本，使用该样本初始化跟踪模型；

[0036] 初始化完成后；以前一帧目标位置为均值，以$(0.09r^2,0.09r^2,0.25)$为协方差，产生候选样本，其中：r为前一帧目标框的宽和高的平均值。

[0037] 作为本发明进一步的方案：所述步骤S22包括；

[0038] 将候选样本分别输入至第一卷积层模块、第一多挑战感知器模块中，第一卷积层模块用来提取通用的目标特征，第一多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第一多挑战感知器模块的输出结果通过第一concatenate函数层进行通道维度的拼接，并输送至第一聚合层模块，第一聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第一聚合层模块处理的结果与第一卷积层模块提取的目标特征进行相加融合，输送至第二卷积层模块以及第二多挑战感知器模块处，依次传递，直至到达Adaptive RoI Align层，Adaptive RoI Align层加快跟踪过程中候选区域的特征提取过程，根据不同候选样本提取对应位置的深度特征得到最终的特征图。

[0039] 作为本发明进一步的方案：所述步骤S23包括；最终的特征图输入分类器模块中，通过FC6获得每个候选样本被判定为正样本和负样本的得分，分别设为$f^+(x^i)$和$f^-(x^i)$，利

用公式$x^{*}=\arg\max_{x^{i}} f^{+}(x^{i})$确定当前帧的目标位置，其中$x^{i}$表示采样的第i个样本，$f^{+}(x^{i})$表示获取的正样本得分，$f^{-}(x^{i})$表示获取的负样本得分；$x^{*}$为预测的目标位置。

[0040] 作为本发明进一步的方案：所述步骤1)为；跟踪成功时，在当前帧的预测位置周围采集正样本和负样本，将这些样本以Adaptive RoI Align后的特征保存至总的正负样本数据集中；

[0041] 所述步骤2)为；跟踪失败时，进行短期更新，短期更新包括：从总正负样本数据集中抽出最近20帧跟踪成功收集的正负样本进行迭代训练；设定FC4、FC5的学习率为0.0003，FC6的学习率为0.003，batchsize为128，其中包含32个正样本和96个负样本，共迭代15次，微调FC4、FC5、FC6的权重参数。

[0042] 作为本发明进一步的方案：所述步骤S24还包括长期更新，其更新规则为固定每隔若干帧执行更新。长期更新规则为规则为固定每隔10帧进行更新一次；在对最终跟踪模型进行更新后，判断当前帧是否为最后一帧，若是最后一帧，目标识别及跟踪结束，否则跟踪继续。

[0043] 本发明的优点在于：

[0044] 1、本发明中的模型由依次串联的多级挑战模块、Adaptive RoI Align层、分类器模块组成；多级挑战模块能够学习不同层次的挑战特征，引入Adaptive RoI Align层来加快跟踪过程中候选区域的特征提取过程；有效的增加了特征表达的丰富性，提高了跟踪的鲁棒性，并达到了实时的跟踪性能。

[0045] 2、本发明的多级挑战模块中，引入多层次与主干网络模块并行的多挑战感知器模块学习不同层次的挑战特征，引入聚合层模块来解决跟踪过程中挑战不可知的问题，引入Adaptive RoI Align层来加快跟踪过程中候选区域的特征提取过程；有效的增加了特征表达的丰富性，提高了跟踪的鲁棒性，并达到了实时的跟踪性能。

[0046] 3、本发明并行的多挑战感知器模块能够很好的提取多层次的挑战信息，从而使得挑战属性的特征表达足够丰富，同时保证特征的准确性。

附图说明

[0047] 图1为网络模型结构示意图。

[0048] 图2为本发明中实施例1的流程方框图。

[0049] 图3为本发明中实施例2的流程方框图。

[0050] 图4为是基于多挑战感知学习模型的实时视觉跟踪方法的流程图。

具体实施方式

[0051] 为使本发明实施例的目的、技术方案和优点更加清楚，下面将结合本发明实施例，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。

[0052] 实施例1

[0053] 如图1，图1为网络模型结构示意图；基于多挑战感知学习模型的训练方法，包括以

下步骤；

[0054] S11、构建网络模型；

[0055] 获取当前跟踪视频序列的第一帧，通过给定的第一帧中目标的真值框，以真值框的中心点为均值进行高斯分布采样获取候选样本，本实施例中以$(0.09r^2,0.09r^2,0.25)$为协方差，产生256个候选样本；

[0056] 其中：r为前一帧目标宽和高和的平均值，

[0057] 获取当前跟踪视频序列为现有技术，如通过摄像机等获取，此处不再进行详细描述，同时高斯分布采样也为现有技术。

[0058] 如图2，所述网络模型包括依次串联的用于获取候选样本特征图的多级挑战模块、Adaptive RoI Align层、分类器模块；具体的本实施例中，

[0059] 所述多级挑战模块为第一级挑战模块、所述第二级挑战模块、第三级挑战模块、Adaptive RoI Align层、分类器模块；其中，

[0060] 所述第一级挑战模块包括第一卷积层模块、第一多挑战感知器模块、第一concatenate函数层、第一聚合层模块，将候选样本分别输入至第一卷积层模块、第一多挑战感知器模块中，第一卷积层模块用来提取通用的目标特征，第一多挑战感知器模块提取不同挑战属性下的目标特征表示，包括Scale change(尺度变化)、Target Motion(目标运动)、Camera Motion(相机移动)、Occlusion(遮挡)、Illumination variation(光照变化)，然后第一多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接，并输送至第一聚合层模块，第一聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第一聚合层模块处理的结果与第一卷积层模块提取的目标特征进行相加融合，输送至所述第二级挑战模块处；

[0061] 图2中的Scale(即为Scale change)、Target Motion、Camera Motion、Occlusion、Illumination(即Illumination variation)即为第一多挑战感知器模块提取不同挑战属性下的目标特征表示。

[0062] 所述所述第二级挑战模块包括第二卷积层模块、第二多挑战感知器模块、第二concatenate函数层、第二聚合层模块，所述第二卷积层模块、第二多挑战感知器模块接收到第一级挑战模块输出的相加融合的结果；第二多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第二多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接，并输送至第二聚合层模块，第二聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第二聚合层模块处理的结果与第二卷积层模块提取的目标特征进行相加融合；输送至第三级挑战模块处；

[0063] 所述第三级挑战模块包括第三卷积层模块、第三多挑战感知器模块、第三concatenate函数层、第三聚合层模块，所述第三卷积层模块、第三多挑战感知器模块接收到所述第二级挑战模块输出的相加融合的结果；第三多挑战感知器模块提取不同挑战属性下的目标特征表示，然后第三多挑战感知器模块的输出结果通过concatenate函数层进行通道维度的拼接，并输送至第三聚合层模块，第三聚合层模块将得到的多挑战特征进行聚合处理解决跟踪过程中挑战不可知的问题，第三聚合层模块处理的结果与第三卷积层模块提取的目标特征进行相加融合；输送到达Adaptive RoI Align层，Adaptive RoI Align层加快跟踪过程中候选区域的特征提取过程，根据不同候选样本提取对应位置的深度特征得

到最终的特征图，再输送至分类器模块，分类器模块进行计算得到每个候选样本的得分，最后取最高正样本得分的样本位置作为当前帧预测的视觉跟踪结果。

[0064] 优选的，所述第一卷积层模块、第二卷积层模块、第三卷积层模块作为主干网络模块，通过使用ImageNet数据集进行预训练分类网络VGG-M，并将这个网络的原有参数作为主干网络模块的初始化参数；

[0065] 且第一卷积层模块由依次串联的第一卷积层、ReLU(Rectified Linear Unit，线性整流函数)激活函数层、局部响应归一化层(LRN，Local Response Normalization)、池化核尺寸为3*3的最大池化层(max pooloing)组成；

[0066] 所述第二卷积层模块由依次串联的第二卷积层、ReLU激活函数层、局部响应归一化层(LRN)组成；

[0067] 第三卷积层模块由依次串联的第三卷积层、ReLU激活函数层组成。

[0068] 进一步的，本实施例中，其中第一卷积层、第二卷积层、第三卷积层的卷积核大小分别为7*7、5*5、3*3，第一卷积层、第二卷积层操作步长为2，第三卷积层是操作步长为1、空洞率为3的空洞卷积。

[0069] 所述多挑战感知器模块由依次串联的卷积层、ReLU激活函数层、批归一化层、随机失活层、最大池化层组成。

[0070] 其中，所述聚合层模块由依次串联的卷积层、ReLU激活函数层、局部响应归一化层组成。

[0071] 同时，所述分类器模块是由依次串联的两个全连接层和一个带由softmax层的全连接层所组成。

[0072] 本实施例中，两个全连接层为FC(fully connected，全连接层)4、FC5，且所述FC4、FC5带有随机失活层(图中未画出)和ReLU激活函数层(图中未画出)。

[0073] S12、使用标定好目标的VOT数据集来训练整个所述的网络模型；训练过程包括；

[0074] S1201、在第一帧中根据给定的真值框选取$S_+=50$(IOU≥0.7)和$S_-=200$(IOU≤0.5)的样本数；其中，S_+表示正样本，S_-表示负样本，IOU表示采集样本与真值框之间的交并比；

[0075] S1202、通过采集的正负样本，使用随机梯度下降法进行1000次迭代训练，每次迭代训练根据以下方法处理：设K表示VOT数据集中的视频序列的个数(K为正整数)，为每个视频序列构建一个新的随机初始化的FC6全连接层。

[0076] 需要说明的是，随机梯度下降法以及构建新的FC6全连接层为现有技术，此处不再详细说明。

[0077] 在迭代训练过程中，每次迭代都需要保证每个全连接层使用的是与其对应的视频序列来进行训练，因此在第x轮迭代中的minibatch(batchsize(批尺寸)=128)是从(x mod K)第个视频序列中随机抽取8帧图像通过高斯分布采集正负样本产生，其中包含32个正样本和96个负样本，使用对应的全连接层计算每个样本的前景和背景得分；

[0078] 其中，mod代表求余函数；

[0079] S1203、训练分为两个阶段，第一阶段提取VOT数据集中每个视频序列不同挑战帧的集合，用这些数据训练多挑战感知器模块，其中每个多挑战感知器模块都是单独训练的，主干网络模型的初始化参数为在ImageNet数据集上预训练的VGG-M前三个卷积层的参数，

FC4,FC5随机初始化,在这个阶段中主干网络模块的参数保持不动,每个多挑战感知器模块的学习率均为0.0005,FC4,FC5,FC6的学习率为0.0001,训练过程步骤S1202所示,保存每个多挑战感知器模块分支训练的模型,用于第二阶段的训练;其中,FC4,FC5随机初始化为现有技术,此处不再说明。第二阶段使用整个VOT数据集训练聚合层模块,主干网络模块、FC4、FC5参数设定方式和第一阶段训练相同,并行的多挑战感知器模块的参数为第一阶段训练的参数,在这个阶段保持并行的多挑战感知器模块的参数固定不动,主干网络模块、FC4、FC5、FC6的学习率设置均为0.0001,聚合层模块的学习率设置为0.0005,训练过程如步骤S1202所示,第二阶段训练结束后,得到最终的训练模型,用来进行视觉目标跟踪。

[0080] 实施例2

[0081] 如图1、图3、图4,图1为网络模型结构示意图;图3为本发明中实施例2的流程方框图;图4为是基于多挑战感知学习模型的实时视觉跟踪方法的流程图;

[0082] 基于多挑战感知学习模型的实时视觉跟踪方法,包括以下步骤;

[0083] S21、输入当前跟踪的视频帧,在前一帧预测的目标位置周围用高斯采样获取当前帧的候选样本;

[0084] 由待跟踪视频序列的提供的第一帧图像作为前一帧;由前一帧和框定目标位置区域的真值框,按照高斯分布随机产生5500个样本,$S_+=500$(IOU≥0.7)和$S_-=5000$(IOU≤0.3);

[0085] 使用5500个样本初始化跟踪模型,将这些样本设为batchsize=128大小的minibatch进行初始化训练,构建新的FC6层;其中包含32个正样本和96个负样本;

[0086] 初始化过程中,固定第一卷积层、第二卷积层、第三卷积层的参数,设定FC6层的学习率为0.001,FC4、FC5学习率设定为0.0005,共迭代50次,完成初始化;

[0087] 初始化完成后;以前一帧的目标位置为均值,以$(0.09r^2,0.09r^2,0.25)$为协方差,产生256对候选样本,其中:r为前一帧目标框的宽和高的平均值;

[0088] S22、获取候选样本的的特征图;

[0089] 将候选样本送入到主干网络模块和与其并行的多挑战感知器模块中,依次传递至Adaptive RoI Align层中,根据不同候选样本提取对应位置的深度特征得到最终的特征图;

[0090] S23、将所述特征图输入到分类器模块中,预测目标位置;

[0091] 通过FC6获得每个候选样本被判定为正样本和负样本的得分,分别设为$f^+(x^i)$和$f^-(x^i)$,而利用公式$x^*=\arg\max_{x^i} f^+(x^i)$确定当前帧的目标位置,其中$x^i$表示采样的第i个样本,$f^+(x^i)$表示获取的正样本得分,$f^-(x^i)$表示获取的负样本得分;$x^*$为预测的目标位置。

[0092] S24、判断当前帧是否跟踪成功,正样本的得分大于0时,跟踪成功,执行步骤1);正样本的得分小于0时,即跟踪失败,执行步骤2);

[0093] 所述步骤1)为:在当前帧的预测位置周围采集50个正样本(IOU≥0.6)和200个负样本(IOU≤0.3),并加入总的正负样本数据集中,由于第一卷积层、第二卷积层、第三卷积层和Adaptive RoI Align层的参数在跟踪过程中参数不变,故我们可将这些样本以Adaptive RoI Align后的特征保存至总的正负样本数据集中;用于进行模型更新。

[0094] 其中,本实施例中,总的正样本集保存最近100次跟踪成功帧的正样本,总的负样

本集保存最近20次跟踪成功帧的负样本。

[0095] 所述步骤2)为:并进行短期更新,短期更新为:从总的正负样本数据集中抽出最近20帧跟踪成功收集的正负样本进行迭代训练;设定FC4、FC5的学习率为0.0003,FC6的学习率为0.003,batchsize为128,其中包含32个正样本和96个负样本,共迭代15次,微调FC4、FC5、FC6的权重参数。

[0096] 值得注意的是,本发明中,在整个跟踪过程中会设定长期更新,其更新规则为固定每隔若干帧执行更新;

[0097] 本实施例中,固定每隔10帧进行长期更新一次;长期更新为:收集训练样本进行迭代训练,训练样本是由最近100帧成功跟踪收集的正样本和最近20帧成功跟踪收集的负样本所组成;同样设定前两个全连接层的学习率为0.0003,最后一个全连接层的学习率为0.003,batchsize为128,其中包含32个正样本和96个负样本,共迭代15次,微调全连接层的权重参数。

[0098] 在对最终跟踪模型进行更新后,判断当前帧是否为最后一帧,若是最后一帧,目标识别及跟踪结束,否则跟踪继续。

[0099] 如下表1和表2,,表1和表2是本发明的实验结果图,分别在公开的数据集UAV-Traffic和GOT-10K上进行了测试,并将测试结果与其他的跟踪器在SR(成功率)、PR(准确度)和AO(平均重叠率)上进行了评估。其中HCAT表示本发明的跟踪结果精度,可以很明显的看到相比于其他方法,其跟踪性能均匀较大程度的提升,此外本发明的跟踪方法还可以达到实时的跟踪性能(29fps),对跟踪任务来说有着重要的意义。

[0100] 表1和表2中,SINT为Siamese instance search for tracking;

[0101] HDT:Hedged deep tracking;

[0102] CCOT:Beyond correlation filters:Learning continuous convolution operators for visual tracking;

[0103] CFNet:End-to-end representation learning for correlation filter based tracking;

[0104] SiamFC:Fully-convolutional siamese networks for object tracking.

[0105] ECO:Eco:Efficient convolution operators for tracking;

[0106] RT-MDNet:Real-time mdnet.

[0107] MDNet:Learning multi-domain convolutional neural networks for visual tracking.

[0108] ANT:Learning attribute-specific representations for visual tracking.

[0109] HCAT:Learning Hierarchical Challenge-Aware Representations for Real-Time Visual Tracking(即本发明的方法)

[0110]

Trackers	SINT	HDT	CCOT	CFNet	SiamFC	ECO	RT-MDNet	MDNet	ANT	HCAT
PR	57.0	59.6	65.9	68.0	68.1	70.2	71.4	72.5	77.0	77.5
SR	29.0	30.3	40.9	42.8	44.7	45.1	44.2	46.4	46.3	46.4

[0111] 表1

[0112]

Method	AO	$SO_{0.50}$
RT-MDNet(Jung et al. 2018)	0.341	0.358
SiamFC(Bertinetto et al. 2016)	0.348	0.353
GOTURN(Held, Thrun, and Savarese 2016)	0.347	0.375
CCOT(Danelljan et al. 2016)	0.325	0.328
ECO(Danelljan et al. 2017)	0.316	0.309
MEEM(Zhang, Ma, and Sclaroff 2014)	0.253	0.235
HCAT	**0.379**	**0.404**

[0113] 表2

[0114] 以上实施例仅用以说明本发明的技术方案，而非对其限制；尽管参照前述实施例对本发明进行了详细的说明，本领域的普通技术人员应当理解：其依然可以对前述各实施例所记载的技术方案进行修改，或者对其中部分技术特征进行等同替换；而这些修改或者替换，并不使相应技术方案的本质脱离本发明各实施例技术方案的精神和范围。

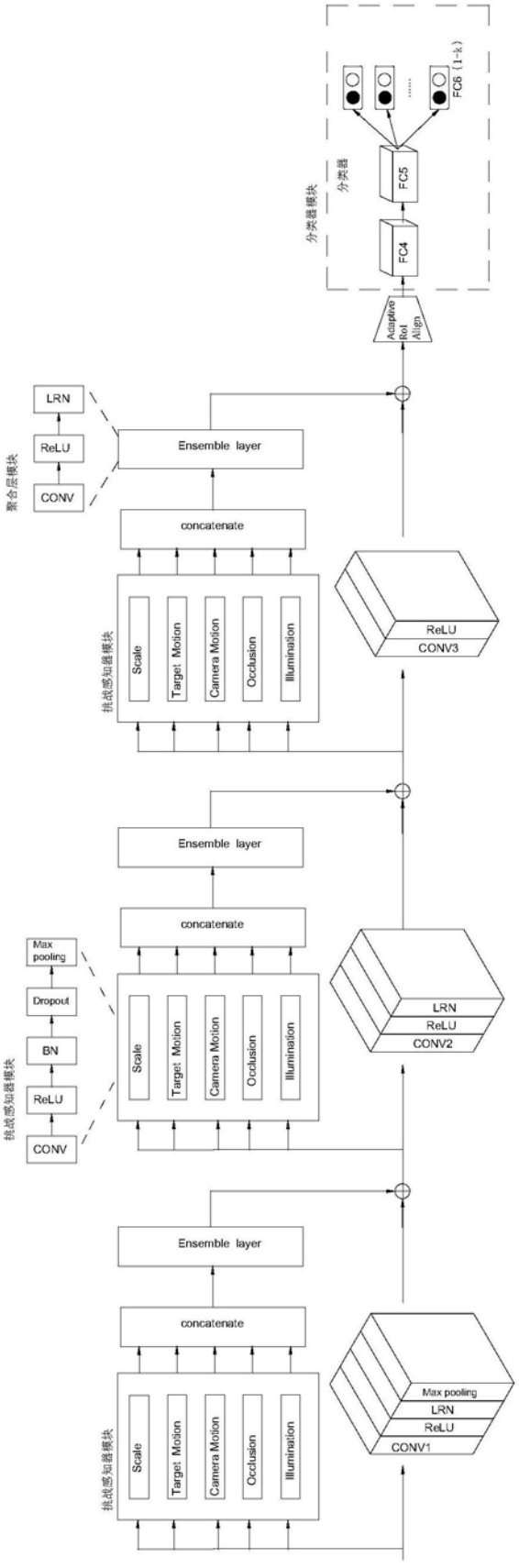

分类器模块
分类器
FC6 (1+k)
FC5
FC4
Adaptive RoI Align
聚合层模块
LRN
ReLU
CONV
Ensemble layer
concatenate
挑战感知器模块
Scale
Target Motion
Camera Motion
Occlusion
Illumination
ReLU
CONV3
Ensemble layer
concatenate
挑战感知器模块
Max pooling
Dropout
BN
ReLU
CONV
Scale
Target Motion
Camera Motion
Occlusion
Illumination
LRN
ReLU
CONV2
Ensemble layer
concatenate
挑战感知器模块
Scale
Target Motion
Camera Motion
Occlusion
Illumination
Max pooling
LRN
ReLU
CONV1

图1

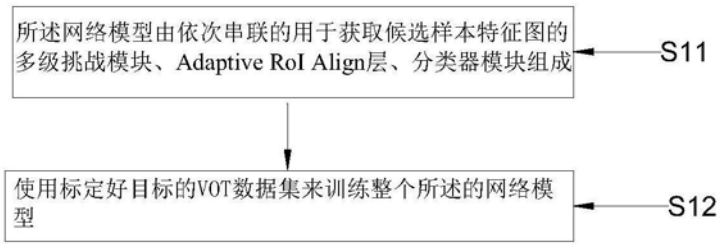

所述网络模型由依次串联的用于获取候选样本特征图的多级挑战模块、Adaptive RoI Align层、分类器模块组成
S11
使用标定好目标的VOT数据集来训练整个所述的网络模型
S12

图2

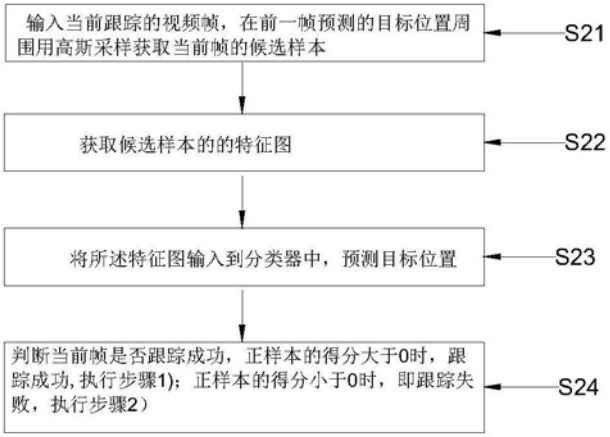

输入当前跟踪的视频帧，在前一帧预测的目标位置周围用高斯采样获取当前帧的候选样本
S21
获取候选样本的的特征图
S22
将所述特征图输入到分类器中，预测目标位置
S23
判断当前帧是否跟踪成功，正样本的得分大于0时，跟踪成功,执行步骤1)；正样本的得分小于0时，即跟踪失败，执行步骤2）
S24

图3

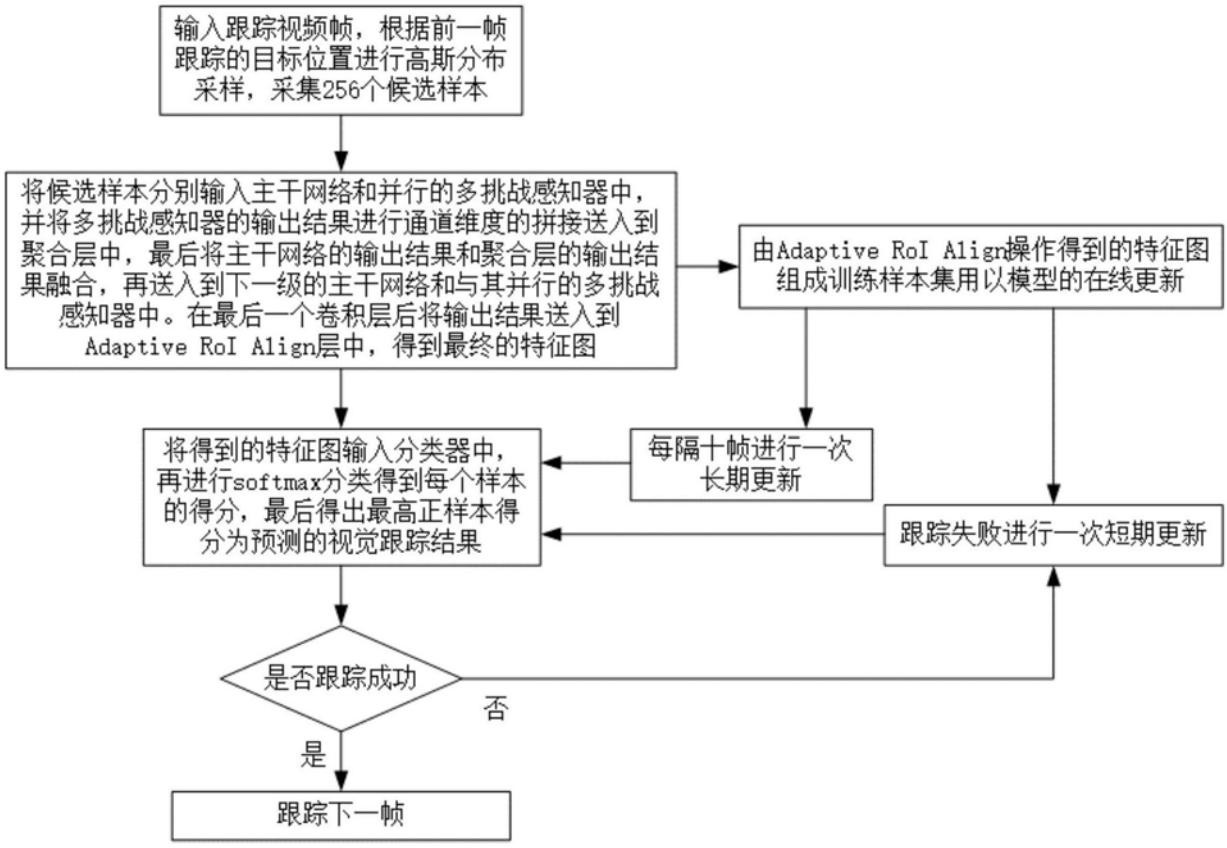

输入跟踪视频帧，根据前一帧跟踪的目标位置进行高斯分布采样，采集256个候选样本
将候选样本分别输入主干网络和并行的多挑战感知器中，并将多挑战感知器的输出结果进行通道维度的拼接送入到聚合层中，最后将主干网络的输出结果和聚合层的输出结果融合，再送入到下一级的主干网络和与其并行的多挑战感知器中。在最后一个卷积层后将输出结果送入到Adaptive RoI Align层中，得到最终的特征图
由Adaptive RoI Align操作得到的特征图组成训练样本集用以模型的在线更新
将得到的特征图输入分类器中，再进行softmax分类得到每个样本的得分，最后得出最高正样本得分为预测的视觉跟踪结果
每隔十帧进行一次长期更新
跟踪失败进行一次短期更新
是否跟踪成功
否
是
跟踪下一帧

图4