

融合时空上下文信息的强化学习小目标快速搜索

姜 虹¹, 马姣姣¹, 姚红革¹, 程嗣怡², 陈 游², 喻 钧¹

(1. 西安工业大学计算机科学与工程学院, 陕西西安 710021; 2. 空军工程大学航空工程学院, 陕西西安 710038)

摘 要: 人眼在搜索目标时, 先基于此前的扫视经验粗略扫视, 找到可能有目标的位置, 再进行详细搜索. 前者的扫视可称为基于时间上下文信息的扫视, 后者可称为基于位置上下文信息的搜索. 受人眼这种目标搜索模式启发, 本文提出一种结合强化学习的时空上下文目标搜索方法. 该方法基于强化学习搜索策略构建时间上下文模块, 获得时间上下文信息; 再通过构建一个自适应多尺度窗口提取位置上下文信息, 两种信息在目标搜索过程中交替配合, 完成目标搜索. 实验结果表明, 该方法在 MS COCO 数据集上较基准方法提升了 2.9%, 且可在 5 个搜索次数内找到目标.

关键词: 小目标检测; 强化学习; 时间上下文; 位置上下文; 人眼扫视

中图分类号: TP391.4

文献标识码: A

文章编号: 0372-2112(2023)11-3176-11

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20220617

Rapid Search for Small Object in Reinforcement Learning by Combining Spatio-Temporal Contextual Information

JIANG Hong¹, MA Jiao-jiao¹, YAO Hong-ge¹, CHENG Si-yi², CHEN You², YU Jun¹

(1. School of Computer Science and Engineering, Xi'an Technological University, Xi'an, Shaanxi 710021, China;

2. Aeronautics Engineering College, Air Force Engineering University, Xi'an, Shaanxi 710038, China)

Abstract: When searching for a object, the human eye first roughly scans based on previous scanning experience to find potential locations for the object, and then conducts a detailed search. The former can be referred to as scanning based on temporal contextual information, while the latter can be referred to as searching based on location contextual information. Inspired by this, this paper proposes a rapid search method for small objects based on reinforcement learning that integrates spatio-temporal context information. The method builds a temporal context module based on a reinforcement learning search strategy to simulate the human eye's ability to obtain and utilize empirical information, then constructs an adaptive multi-scale window to extract location context information to simulate the human eye's ability to search carefully at possible locations. The two kinds of information cooperate alternately in the object search process to complete the object search. The experimental results show that the proposed algorithm brings around 2.9% gain on MS COCO benchmark, and can find an object within five search counts.

Key words: small object detection; reinforcement learning; temporal context; location context; human eye saccade

1 引言

人眼搜索目标的过程可以总结为两次扫视^[1]: 首先大致扫视整个场景, 随后在具有可能性大的位置上再一次扫视. 前一步可以称“粗扫”, 后一步为“细扫”. 这种从整体到局部的“扫视”对目标的尺寸并不敏感, 这为本文探索如何模拟人眼的搜索模式来提高小目标的检测准确度提供了启发. 此外, 这种从外到内、从粗到细的“扫视”还涉及目标周围有关的环境信息和一些包含历史信息的先验知识, 本文称之为时空上下文(位置上下文和时间上下文), 这两种上下文为搜索目标提供

了必要的信息.

受上述人眼搜索模式的启发, 本文提出了一个基于位置和时间上下文的强化学习搜索方法来解决小目标检测问题. 该方法通过一个循环结构来模拟人眼扫视过程, 并结合强化学习的序列决策策略^[2], 在时间序列上利用包含历史信息的“时间上下文”在含有位置信息的“位置上下文”上执行一系列动作来缩小寻找目标的范围.

首先, 构建一个自适应多尺度窗口提取模块, 来获取兼顾了低冗余度和高信息量的位置上下文窗口, 这

些窗口包含小目标信息,并且可随着搜索范围的变换而变换.其次,为了增强小目标的特征表示,本文在文献[3,4]的基础上设计了一个双向特征融合网络,旨在通过语义特征和空间特征的双向传递来增强小目标的特征表示.再次,模拟人眼从整体到局部的搜索过程,设计一个包含历史信息的时间上下文搜索模块,利用该时间上下文逐渐缩小搜索范围,将注意力最终聚焦在小目标上.最后,设计一个判别网络,以判断提取的窗口是否包含完整目标以及是否终止搜索.

这种搜索方式强调的是在出现目标可能性大的区域上进行搜索,所以比起主流目标检测框架需要的数千个 anchors,它只需少量的证据区域就能定位到目标.如图1所示,黄框为证据窗口($Cr=1$ 中置信度最大的窗口),绿框为执行动作后的可视区域,红框表示最终搜索到的目标边界框.图1中右侧是对左侧目标搜索行为过程的说明.在 t_i 时刻,将整个图像划分为若干个窗口,并判断它们是否包含完整的目标.若没有,则代理根据证据窗口以及历史信息(时间上下文)执行动作,将产生相应的可视区域送入下一时刻进行下一轮循环,直到搜索到目标.

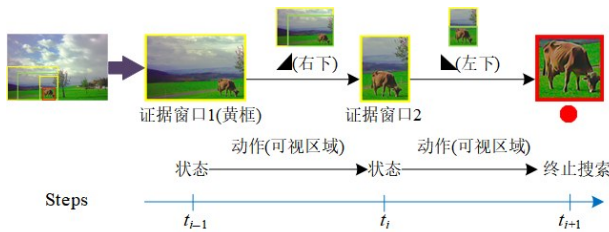


图1 搜索目标采取的一系列操作

本文的主要贡献如下.

(1)设计了一个自适应多尺度窗口提取模块,它能获取低冗余度和高信息量的含有位置上下文信息的窗口.定义了一个新的评价指标即完备率 Cr (Completeness ratio),以衡量多尺度窗口中目标的完整性.

(2)设计了一个双向特征融合网络,旨在融合用于细节特征的语义信息和用于定位的空间信息来增强小目标的特征表示.

(3)模拟人眼从整体到局部的搜索方式,结合强化学习设计一个时间上下文搜索模块.它能够利用历史(经验信息)在位置上下文中一步一步地缩小搜索范围,最终将注意力聚焦在小目标上.

基于上述3点,本文构建了一个具有时空上下文搜索能力的强化学习小目标快速检测网络 St-CISNet (Spatio-temporal Contextual Information Search Network).在PASCAL VOC^[5]和MS COCO^[6]数据集上进行了实验,并将实验结果与目前先进的检测算法进行了比较.实验表明,St-CISNet指导下的代理能够在分析图

像中的3~5个区域后定位目标的单个实例,并且针对图像中的多个目标,在不使用区域建议框的情况下能获得最佳检测结果.

2 相关工作

2.1 基于层信息融合的目标检测

小目标经过深层卷积后会丧失较多信息,而一些经典的算法通常在最高层的特征图上检测目标^[7,8],这导致小目标检测精度难以提升.针对这一点,Liu等^[9]利用感受野小的浅层特征检测对分辨率敏感的较小目标,而利用感受野大的深层特征检测对分辨率不敏感的较大目标.之后,许多研究工作^[10,11]也遵循这一原理来检测多尺度目标.Lin等^[12]发现浅层特征语义信息较少但目标位置准确,深层特征语义信息丰富但目标位置粗略,针对此现象,提出一种具有侧向连接的自顶向下的特征金字塔网络,称为FPN (Feature Pyramid Network).FPN通过将相邻层的特征融合,达到小目标特征增强的目的.但是FPN仅仅传递了语义信息,并没有对细节信息进行传递.Liu等^[3]为了融合细节信息提出一种路径聚合网络PANet (Path Aggregation Network),将自底向上的增强路径和自适应特征池化结合,在多个数据集上达到最先进的表现.最近,Leng等^[4]提出了一种类似于PANet的融合方法Bi-FFM (Bidirectional Feature Fusion Module),通过在神经网络中进行两次传递来增强小目标的特征表示.然而,这些算法只增强了用于细节特征的语义信息,忽略了用于定位的空间信息.

2.2 基于上下文学习的目标检测

将上下文作为额外信息以辅助目标检测是近年来研究的一大热点^[13,14].Tang等^[15]提出了一种基于锚的显式上下文辅助方法,用于在学习小人脸的上下文特征中引入监督信息.为了进一步提升目标的分类结果,Pato等^[16]通过使用图像中所有检测结果的非视觉高级上下文信息,来对先前检测器的检测结果进行上下文重打分.此外,针对小目标的误检和漏检问题,Fu等^[17]将目标之间的内在语义和空间布局关系作为上下文进行建模和推断,在提取小目标内在语义的同时尽可能保留其空间信息.最近,Lim等^[18]也提出了一种利用隐式上下文的目标检测方法.它主要考虑位于目标周围的背景信息,通过利用这些背景信息来推测目标.与它类似,Leng等^[4]提出的IENet (Internal-External Network)也考虑到与目标有联系的背景信息.它充分利用内部特征信息和外部环境信息来提高小目标的检测精度.尽管上下文在目标检测中已经得到充分利用,但是现有的方法还是无法将显式上下文(位置特征、空间信息)与隐式上下文(背景特征、时间信息)结合起来用于小目标检测.

2.3 基于强化学习的目标检测

基于强化学习的目标检测方法^[19]不同于基于深度卷积网络的目标检测方法^[20,21],它通常将目标检测问题转换为一个智能体搜索目标的序列决策问题. Mnih等^[19]为了模拟人眼的视觉注意机制提出了循环注意力模型,成功地将强化学习应用到了目标识别领域. Caicedo等^[22]提出了一种基于DQN(Deep Q-Learning)^[2]的主动目标定位方法,通过自上而下的推理使智能体将注意力集中在候选区域,从而确定目标的最终位置. Jie等^[23]提出了一种树结构深度强化学习方法. 该方法能够充分利用当前观测路径和历史搜索路径来顺序搜索目标. 尽管强化学习和深度学习结合得很好,但是他们的识别精度并不高,其中一个主要原因是现有的基于强化学习的目标检测方法通常将最后一步非最优区域作为检测结果. 基于此,最近Zhou等^[24]将强化学习代理的动作空间与基于卷积神经网络的特征空间相结合,通过强化区域的选择和细化能力有效地提升了目标检测的准确度. 但是该方法在小目标上的结果差强人意,其主要原因是:(1)对于复杂场景下的小目标,特征往往难以提取;(2)仅仅依赖局部信息来识别小目标时具有较大的局限性.

3 St-CISNet模型

本文提出的St-CISNet模型框架如图2所示. 该模型是一个循环结构,主要包含3个模块:自适应多尺度窗口提取模块、上下文判别模块和时间上下文搜索模块(分别简称为提取模块、判别模块、搜索模块). 其中,提取模块获取低冗余度和高信息量的含有位置上下文信息的窗口;判别模块判断提取的窗口是否包含完整目标以及是否终止搜索;搜索模块结合强化学习,利用时间上下文逐渐缩小搜索范围,以将注意力最终聚焦在小目标上. 整个搜索中的决策功能主要靠强化学习来完成,时空上下文承担的是搜索中的信息收集功能.

St-CISNet首先利用提取模块将输入图像划分为若

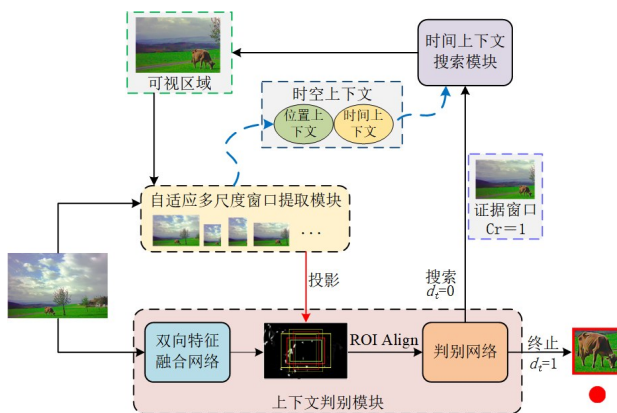


图2 St-CISNet模型框架

干个包含位置上下文信息的多尺度窗口(位置上下文);同时,利用双向特征融合网络,从原图中提取具有语义信息和位置信息的特征. 其次,将多尺度窗口按对应的尺寸比例投影到特征图上获得相应的多尺度窗口特征矩阵,再通过RoI Align层缩放到 7×7 大小,送入判别网络. 若多尺度窗口的置信度大于 $0.5(d_t=1)$,则表示找到目标,终止搜索;否则,将 $Cr=1$ 的窗口作为证据窗口,并从证据窗口中选择置信度最大的窗口送入搜索模块,将历史信息 and 当下观察作为状态输入(时间上下文),依据从动作空间中选择的具有最大折扣预期奖励的动作在包含位置上下文的证据窗口进行搜索. 同时,将搜索到的新可视区域再次送入提取模块进行下一轮循环,直到搜索到目标或者因搜索步长过长而终止搜索.

3.1 自适应多尺度窗口提取模块

该模块的主要目的是模仿人眼扫视到的位置信息,利用多尺度矩形框提取包含位置上下文信息的窗口. 具体的提取方法是:以与原图的3种比例(2:1, 5:3, 5:4)生成3种大小的矩形框,并将这3种大小的矩形框旋转 180° ,得到6种不同尺寸的矩形框,如图3(a)所示. 每种尺寸的矩形框将分别在原图的中心、左上、左下、右上、右下5个位置进行多尺度窗口选取,如图3(b)所示. 对于一个大小为 $W \times H$ 的图像,一次循环可以提取到30个多尺度窗口.

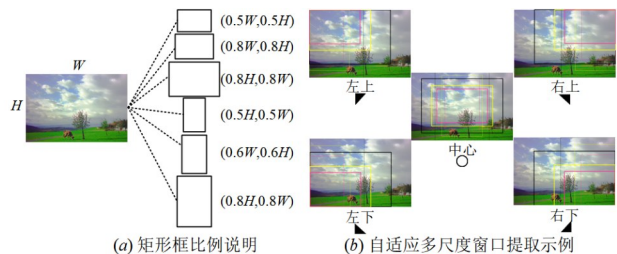


图3 自适应多尺度窗口提取示意图

这种提取方式,一是考虑到目标的尺寸,0.6与0.8的提取比例会尽可能地框选较大目标,0.5的提取比例不会遗漏图像信息且可以减少对较小目标的搜索次数;二是考虑到一些目标的长宽比与原图长宽比相反,旋转后的矩形框可以更好地框选这类目标(详见5.2.1节的可视化结果);三是考虑到目标在原图中的位置,无论目标处于原图何处,都可从这5个位置来把握目标. 本提取方法有以下两方面优势.

一是在获取小目标信息的同时,为搜索小目标提供更多的位置上下文信息. 通过多尺度的矩形框可以获取到丰富的小目标信息,同时得到包含目标位置上下文信息的窗口. 这些窗口为搜索目标提供重要的证据信息,以引导模型寻找最终目标.

二是这种提取方法兼顾了低冗余度和高信息量. 与基于 Anchor 的目标检测算法不同, 本文强调的是在出现目标可能性大的区域上进行搜索, 并不是直接将数千个 Anchors 与真值框进行匹配. 所以仅需要提供给搜索足够的证据区域即可. 实验表明, 3 种比例、1 种变换在 5 个位置框选目标既不会遗漏, 也不会造成信息冗余. 比起滑动窗口^[25], selective search 算法^[26]以及区域建议网络 (Region Proposal Network, RPN)^[6]提取到 2 000 多个候选区域, 多尺度矩形框的提取方法显然大大减少了信息的冗余. 特别是, 这种提取方式基于可视区域, 当可视区域发生变化时, 窗口也能进行自适应, 这种灵活性对小目标更加友好.

3.2 上下文判别模块

该判别模块判断提取的窗口是否包含完整目标. 它包含双向特征融合网络和判别网络. 其中, 双向特征融合网络旨在通过语义信息和空间信息的双向传递增强多尺度窗口的特征表示, 而判别网络用于提供搜索和停止的信号, 如图 2 所示. 这两个部分共同建立了一个鉴别性的网络: 首先, 利用嵌入了双向特征融合的特征提取器从原图中提取到相应的特征图; 其次, 将多尺度窗口投影到特征图上获得相应的特征矩阵; 最后, 将每个特征矩阵通过 RoI (Region of Interest) Align 层^[27]缩放到 7×7 大小, 送入基于深度学习的分类器中, 判断是否找到目标并停止搜索.

3.2.1 双向特征融合网络

对于卷积网络来说, 低分辨率的深层特征具有强语义信息, 而高分辨率的浅层特征具有用于定位的空间信息. 自顶向下的特征融合通常是将深层的强语义信息传递到浅层, 只增强了语义特征, 并没有对空间信息进行传递, 这对小目标检测来说显然是不友好的. 本文针对这一点在 PANet^[3]与 IENet^[4]的基础上提出了双向特征融合方法, 旨在通过空间信息的传递进一步增强小目标的特征表示. 这种融合方式也适用于其他网络.

本文采用的主干网络是 VGG16^[28]. 它有 5 个 maxpooling 层, 使用其中的 4 个 maxpooling 层进行融合. 如图 4 所示, 双向特征融合共分为 3 步: 第一步是自顶向下的传递, 将深层的强语义信息向浅层传递 (S_4, S_3 和 S_2 的特征融入 S_1); 第二步是自底向上的传递, 将浅层的空间信息向深层传递 (S_1, S_2 和 S_3 的特征融入 S_4); 第三步是融合两方向的特征得到 C_1, C_2, C_3, C_4 , 用于提取多尺度窗口特征. 最后, 将多尺度窗口按对应的尺寸比例投影到 $C_1 \sim C_4$ 上, 得到多尺度窗口特征矩阵, 将它们通过 RoI Align 层缩放到 7×7 大小, 并送入基于深度学习的分类器中, 判断是否停止搜索.

融合的具体细节: 首先, 自顶向下采用的上采样方

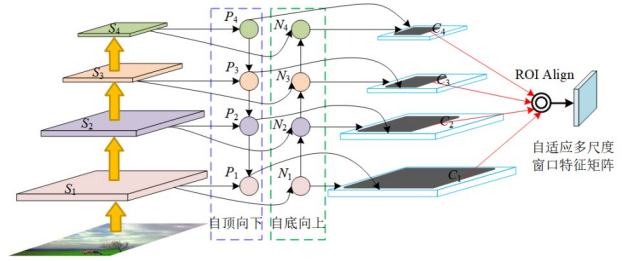


图 4 双向特征融合网络

法是最近邻插值法, 可以最大程度地保留利于分类的语义信息; 其次, 通过 1×1 卷积完成通道数的统一; 最后, 考虑到每一层特征图的不同分布, 利用批处理化来统一每一层特征值的分布^[4].

3.2.2 判别网络

如图 5 所示, 将 7×7 大小的多尺度窗口特征矩阵送入基于深度学习的分类器中. 若多尺度窗口的置信度大于 $0.5 (d_t = 1)$, 则表示找到目标, 终止搜索, 将最大置信度的多尺度窗口作为最终的检测结果; 否则, 将 $Cr = 1$ 的窗口作为证据窗口, 并从证据窗口中选择置信度最大的窗口送入时间上下文搜索模块, 进行下一步搜索.

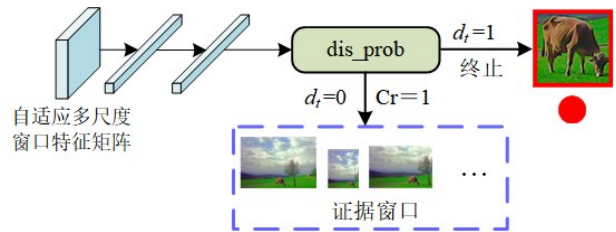


图 5 判别网络

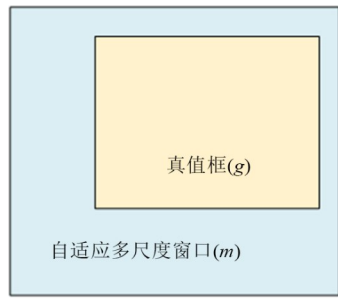
为了确保多尺度窗口中目标的完整性, 本文定义了一个新的评价指标: 完备率 Cr (Completeness ratio). 假设有 C 个样本, 图像中每一个多尺度窗口 $m_i, i \in \{1, 2, \dots, N\}$ 和真值框 $g_j, j \in \{1, 2, \dots, M\}$ 之间的 Cr 定义为

$$Cr = \frac{m_i \cap g_j}{g_j} \quad (1)$$

如图 6 所示, Cr 表示的是多尺度窗口中目标的完整性, $Cr = 1$ 表示多尺度窗口包含了完整的目标.

3.3 时间上下文搜索模块

本模块模仿人眼从整体到局部的搜索方式, 结合强化学习的序列搜索策略, 在时间序列上利用历史信息逐渐缩小搜索范围, 可认为是以时间上下文的形式将注意力最终聚焦在小目标上. 该模块由动作选择网络构成, 如图 7 所示. 利用强化学习 DQN 算法^[2], 将历史信息和当下观察作为状态输入, 依据从动作空间中选择的具有最大折扣预期奖励的动作在包含位置上下文的证据窗口上进行搜索. 同时, 将搜索到的新可视区



$$Cr = \frac{m \cap g}{g} = 1$$

图6 完备率Cr

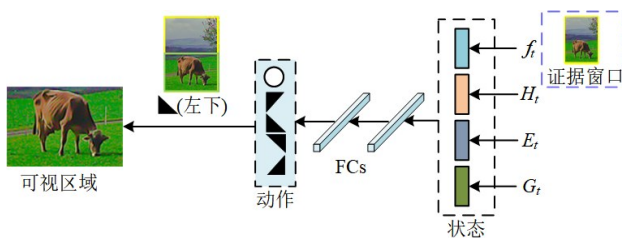


图7 动作选择网络

域再次送入自适应多尺度上下文提取模块进行下一轮循环,直到搜索到目标或者因搜索步长过长而终止搜索. 其中的参数描述如下:

(1)动作. 如图8所示,共有6种可能的动作,其中有5个表示当前证据窗口发生变化的搜索动作 a_G 和1个找到目标或搜索步长过长并结束搜索的终止动作 a_T . 其中,搜索动作是在证据窗口上进行中心、左上、左下、右上和右下的5个操作,这5个动作是基于一个矩形框进行移动的,矩形框的大小设置为证据窗口的0.8倍. 由于本文使用的强化学习算法是DQN,在搜索阶段,本方法一次循环只能处理单个目标. 所以,本文在强化学习的动作参数中增加了一个终止动作,来终止当前的搜索序列,并在一个新的位置重新启动,开始搜索下一个目标. 为了避免一个图像中多个目标对模型搜索特定目标的影响,本文采用IoR(Inhibition of Return)机制^[29]阻止已寻找到的目标再次参与到搜索序列中.



图8 动作说明

(2)状态. 状态是动作历史信息 and 当下观察的图像可视区域信息的融合表示,也称为时间上下文,由元组

$s_t = (H_t, E_t, G_t, f_t)$ 表示. 其中, H_t 是多尺度窗口的历史信息; E_t 是证据窗口的历史信息; G_t 是所采取动作的历史信息; f_t 是当前时刻证据窗口中IoU最大的窗口; s_t 总结了自搜索序列开始以来的多尺度窗口和动作历史.

(3)自适应多尺度窗口历史信息 H_t . 在初始时刻, H_t 包含从原始图像中提取的多尺度窗口 h_t ,之后的每个时间步, H_t 是执行动作后的可视区域 o_t 进入自适应多尺度窗口提取模块产生的新的 h_t . 模型仅对这些窗口进行评估,并决定是否终止搜索. 同时,还使用 H_t 来决定下一个步骤中的证据窗口.

(4)证据窗口的历史信息 E_t . 在上下文判别模块检测完多尺度窗口 h_t 后,将 $Cr=1$ 的窗口作为证据窗口 e_t 保存在 E_t 中. 智能体根据 G_t, H_t, E_t 以及 E_t 中的一个置信度最大的证据窗口 e_t ,共同决定要执行的动作 a_t . e_t 作为上下文搜索当中的重要角色,为搜索目标提供了丰富的位置上下文信息. 将证据窗口 e_t 纳入状态中,也是为了鼓励搜索过程中的多样性,所以每个多尺度窗口最多只能作为证据窗口使用一次. 证据窗口始终从集合 $e_t \setminus E_t$ 中选择:

$$e_t \setminus E_t = \{e_t \in E_t, e_t \notin e_{t-1}\} \quad (2)$$

(5)动作的历史信息 G_t . 每个时间步的可视区域取决于选择的动作历史. 因此,将这段历史纳入状态 s_t 中. 在动作历史向量中,每一个动作都表示一个5维的二进制向量,除了存在的那个动作为1,其余的都为0.

3.4 奖励函数

对于序列决策问题来说,步长也是一种成本. 为了更好地实现小目标检测准确度和计算负载之间的平衡,必须将搜索步长也纳入奖励函数的计算中. 并规定当搜索到目标或者搜索步长达到阈值 $\alpha=10$ 时,都视为终止搜索.

现有的奖励函数如文献[19, 20, 24, 30]提出的奖励函数都只考虑了检测准确率而未考虑到搜索步长,在本文中将采用Mathe等^[31]提出的奖励方案. 当选择动作 a 从状态 s_t 转移到状态 s_{t+1} 时,将奖励函数 r 返回给模型:

$$r = r_a(s_t, s_{t+1}) = \begin{cases} -\beta \cdot |e_t \setminus H_t|, & \text{if } d_t = 0 \\ f_c(h_t) \cdot [\max_{i=1, n} \text{IoU}(h_t, g_i)], & \text{if } d_t = 1 \wedge n > 0 \\ -f_c(h_t), & \text{if } d_t = 1 \wedge n = 0 \end{cases} \quad (3)$$

其中, β 是模型为每个多尺度窗口 h_t 的置信函数评估支付的惩罚, $\beta=10^{-3}$; $f_c(h_t)$ 为 h_t 的置信度, $f_c \in [0, 1]$; d_t 是表示是否终止搜索的信号,若 $d_t=1$,则表示搜索到目标,否则, $d_t=0$;从 E_t 中选择置信度最大的证据窗口 e_t 进行搜索.

奖励函数的第一个分支表示在未能搜索到目标且搜索动作未停止时获得和搜索区域相关的负奖励;第二个分支表示找到目标且搜索步长未超过阈值 $\alpha=10$ 时的正奖励;第三个分支表示在终止信号出现时没有找到目标的负奖励,也就是在搜索步长超过 10 时还没有找到目标.

4 训练

St-CISNet 的参数由双向特征融合网络参数 θ_f 、判别网络参数 θ_d 及动作选择网络参数 θ_a 给出, $\theta = (\theta_f, \theta_d, \theta_a)$. 学习这些参数以最大化折扣累积预期奖励总和.

整个网络的损失分为两部分:一是判别多尺度窗口中是否包含目标且判断正确的判别损失 L_{dis} ,二是预测动作时动作网络的损失 L_{act} .

$$L = L_{\text{dis}} + L_{\text{act}} \quad (4)$$

判别损失由两部分组成:一是分类损失 L_{cls} ,二是边界框回归损失 L_{reg} .

$$\begin{aligned} L_{\text{dis}}(p, u, t^u, v) &= L_{\text{cls}}(p, u) + L_{\text{reg}}(t^u, v) \\ L_{\text{cls}}(p, u) &= -\log p_u \\ L_{\text{reg}}(t^u, v) &= \sum_{i \in \{x, y, w, h\}} \text{smooth}_{L_1}(t_i^u - v_i) \\ \text{smooth}_{L_1}(x) &= \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \end{aligned} \quad (5)$$

其中, p 是分类器预测的 softmax 概率分布; u 对应目标真实类别标签; p_u 是分类器预测当前多尺度窗口为类别 u 的概率; t^u 对应边界框回归器预测的对应类别 u 的回归参数 $(t_x^u, t_y^u, t_w^u, t_h^u)$; v 对应真实目标的边界框回归参数 (v_x, v_y, v_w, v_h) ; (x, y, w, h) 分别表示窗口的中心点坐标和宽高.

动作网络的损失:

$$L_{\text{act}} = (y - Q_{\theta_a}(s, a))^2 - \frac{\lambda}{2} \theta_a^T \theta_a \quad (6)$$

$$y = \begin{cases} r, & \text{terminate} \\ r + \gamma \max_{a'} Q(s', a', \theta_a), & \text{otherwise} \end{cases} \quad (7)$$

其中, (s, a, r, s') 分别是这一时刻的状态 s , 根据状态执行的动作 a , 以及执行动作后产生的奖励 r 和下一时刻的状态 s' ; y 是动作网络的目标值; $Q_{\theta_a}(s, a)$ 是动作网络的预测值; $\lambda/2 \theta_a^T \theta_a$ 是正则化项, 以避免过拟合.

St-CISNet 算法流程如算法 1 所示.

5 实验结果与分析

本文在两个公共目标检测数据集 PASCAL VOC 和 MS COCO 上验证所提 St-CISNet 的有效性. 在与主流的目标检测方法比较的基础上, 也进行了消融试验, 以

算法 1 St-CISNet 算法

输入: 目标训练集与测试集

输出: 目标类别与目标边界框

初始化 $\theta = (\theta_f, \theta_d, \theta_a)$, 状态空间 S , 动作空间 A , 折扣率 γ , 学习率 α

FOR episode=1, M DO

 初始化序列 $s_t = (H_t, E_t, G_t, f_t), h_t \in H_t, e_t \in E_t, a_t \in G_t$,

$f_t = \arg\max_{e_t \in E_t} f_d(h_t)$

 FOR $t=1, T$ DO

 IF $d_t=0$ THEN

 1. 在状态 s_t 选择动作 a_t ,

 2. 执行动作 a_t 得到奖励 $r_a(s_t, s_{t+1})$ 和可视区域 h_t ,

 3. 更新 $H_t, E_t, G_t: H_{t+1} \leftarrow H_t \cup \{h_t\}, E_{t+1} \leftarrow E_t \cup \{e_t\},$

$G_{t+1} \leftarrow G_t \cup \{a_t\}$

 ELSE

 1. 输出目标的检测准确度: $f_d(h_t)$

 2. 通过最小化损失更新模型: $\nabla L = \nabla L_{\text{dis}} + \nabla L_{\text{act}}$

 3. 更新模型参数 $\theta' = (\theta'_f, \theta'_d, \theta'_a)$

 END IF

 END FOR

END FOR

证不同模块对检测精度的影响.

5.1 实验设置

数据集: 在 PASCAL VOC 和 MS COCO 上对 St-CISNet 进行了实验验证, 这两个数据集分别包括 20 类和 80 类目标.

(1) PASCAL VOC. 对于 PASCAL VOC 数据集, 采用 VOC 07+12 的联合训练. 在 VOC 07+12 的 trainval 集上训练模型, 并在包含 4 952 幅图像的 VOC07 test 集上进行评估. 所有模型的主干网络均是 VGG16.

(2) MS COCO. 相比 PASCAL VOC 数据集, MS COCO 数据集中的图像涉及更复杂的场景和更小的目标. 本文在 COCO train2017 集上训练模型, 并在 COCO test-dev 上对其进行评估.

实验细节如下: 对整个模型进行 12 万次迭代, 在对所有训练图像执行一个 episode 后结束, 使用标准正态分布来初始化权重系数, 以 0.002 的初始学习率开始训练, 并针对不同的数据集使用不同的学习率衰减策略. 根据经验, 将每个图像的最大搜索步长设置为 10. 在贪婪训练期间, 前 2 万次迭代中, 从 $\varepsilon=1$ 开始, 以 0.1 为步长递减到 $\varepsilon=0.1$, 并在其余的迭代次数中以 $\varepsilon=0.1$ 进行训练. 将贴现系数设置为 $\gamma=0.9$, 来平衡当前和未来的回报.

5.2 PASCAL VOC 数据集上的性能评估

本节在 VOC 2012 数据集上对 St-CISNet 进行可视化分析, 以直观地说明本文所提方法的有效性, 并在 VOC 2007 的测试集上与主流方法进行比较.

5.2.1 定性实验结果分析

为了验证 St-CISNet 的有效性,本文在 VOC 2012 数据集上进行了实验,以模型在图像中搜索目标轨迹的可视化来展示.如图 9 所示,图中黄框为证据窗口,绿框为执行动作后的可视区域,红框表示最终搜索到的目标边界框.分别从目标所处位置、目标与原图的比例以及目标的尺寸 3 个方面分析模型的有效性.

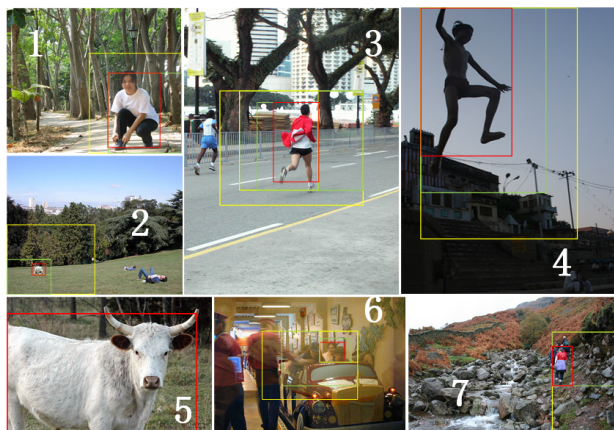


图9 PASCAL VOC 数据集上搜索目标的可视化

(1) 目标所处位置. 图 9 中的 1, 2, 3 号图像仅进行了一两次搜索就找到了目标. 观察图像可以看出,这几幅图像中的目标正好位于整幅图像的右下角、左下角和中心位置,与本文提出的自适应多尺度窗口提取方法恰好对应.所以在搜索时,处于这几个位置的目标会更容易被发现.特别地,如 5 号图像所示,没有进行搜索就已经找到了目标.

(2) 目标与原图的比例. 7 号图像中目标的大小与原图的长宽成反比,一般来说会更难以检测,但从可视化的实验结果来看,St-CISNet 对这种比例关系并不敏感,这主要归功于本文提出的自适应多尺度窗口的提取方法,尤其是 180° 的变换,它决定了这类目标的检测精度.

(3) 目标的尺寸. 2 号和 6 号图像中的目标较小,从可视化结果可以看出,St-CISNet 在一定程度上提高了小目标的识别精度.

此外,为证明 St-CISNet 搜索目标的快速性,本文记录了寻找每个目标的搜索次数,图 10 为部分目标搜索次数的统计.可以看出,St-CISNet 找到目标不会超过 5 次搜索,这大大加快了目标的检测速度.

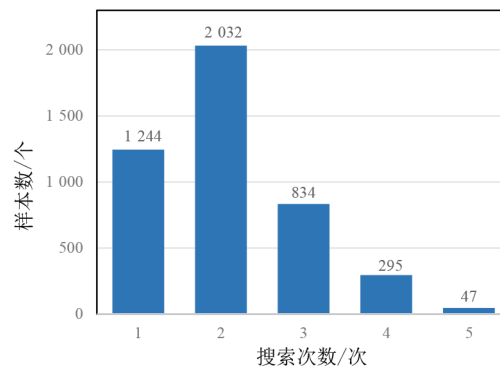


图10 寻找每个目标的搜索次数

5.2.2 与其他方法进行比较

在 VOC 2007 测试集上,将 St-CISNet 与其他方法进行了对比.表 1 展示了不同方法部分目标类别在 VOC 2007 测试集上的检测结果,比较方法包括基于强化学习的 Caicedo-RL^[22], Tree-RL^[23] 和 ReinforceNet^[24],以及单阶段目标检测的典型方法 SSD^[8]、两阶段目标检测的代表方法 Faster R-CNN^[6].与现有的基于强化学习的目标检测方法相比,本文提出的 St-CISNet 方法具有明显的优势,与 ReinforceNet 算法的检测精度相当.相比传统的目标检测算法,检测精度也有明显的提高.并且在候选框数量上,如表 2 所示,本文提出的方法不到它们的 1/10,大大减少了信息的冗余.这主要归功于证据窗口为搜索目标提供的位置上下文信息,以及双向特征融合网络融合了用于定位的空间信息,增强了那些特征不明显目标的特征表示.

表1 不同方法在 VOC 2007 测试集上的检测结果

单位: %

算法	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	mAP
Faster R-CNN ^[6]	76.5	79.0	70.9	65.5	52.1	83.1	84.7	86.4	52.0	81.9	65.7	84.8	84.6	77.5	76.7	73.2
SSD512 ^[8]	82.4	84.7	78.4	73.8	53.2	86.2	87.5	86.0	57.8	83.1	70.2	84.9	85.2	83.9	79.7	76.8
Caicedo-RL ^[22]	55.9	61.9	38.4	36.5	21.4	56.9	58.8	55.9	21.4	40.5	46.3	54.2	56.9	55.9	45.7	46.1
Tree-RL ^[23]	78.2	82.4	78.0	69.3	55.4	86	79.3	88.4	60.8	85.3	74.0	85.7	86.3	78.2	77.2	76.6
ReinforceNet ^[24]	76.5	82.0	74.8	65.2	52.3	80.8	81.1	85.7	56.6	80.3	72.4	83.2	83.4	76.3	77.1	73.7
St-CISNet*	85.3	83.6	78.6	70.3	56.1	82.5	87.6	86.2	61.2	82.1	71.3	84.3	86.9	80.1	78.2	77.4

注: *表示多个随机运行的平均结果,加粗数据为最优结果.

5.3 小目标检测的性能评估

为了检验 St-CISNet 在小目标上的搜索效果,在 MS COCO 数据集上进行了实验.

5.3.1 定性实验结果分析

如图 11 所示,从“杂乱”“光照”“尺度”“遮挡”4 个层面来进行说明.

表2 主流方法的候选框数量

算法	VOC 2007	region proposal number
Faster R-CNN	73.2	≈2k
SSD512	76.8	—
ReinforceNet	73.7	≈50
St-CISNet*	77.4	≈30~150

注:*表示多个随机运行的平均结果.



图 11 MS COCO 数据集上搜索目标的可视化

(1)尺度. 对于尺度变化而言,目标的特征表示是重点,本文提出的自适应多尺度窗口提取方法与双向

特征融合方法刚好满足了这一点,所以 St-CISNet 对于目标尺度的变化并不是很敏感.

(2)遮挡. 遮挡是造成小目标漏检的主要原因之一. 本文并没有使用一些数据增强方法来优化检测效果,但 St-CISNet 对待检测的目标之间的相互遮挡有一定的效果,这主要归功于本文提出的搜索策略以及 IoR 机制避免了一个图像中多个目标对智能体搜索特定目标的影响.

(3)光照. 光照的变化需要检测网络具有一定的鲁棒性. 从实验结果来看,St-CISNet 对光照下的小目标效果并不是非常明显. 主要猜测是光线对模型的影响较大,接下来的工作将利用背景减法对模型进行改进.

(4)杂乱. 在诸多问题中,“杂乱背景”是目标检测中一个极具挑战性的问题. 它包含上述所有难点. 本文所提出的方法在这种特殊的背景下也具有一定的有效性,如图 11 的第一行所示,具有杂乱背景图像的搜索轨迹的可视化.

5.3.2 与其他算法进行比较

表 3 展示了一些基准方法在 MS COCO 测试集上的检测结果. 虽然 MS COCO 数据集更复杂,但是 St-CISNet 在 MS COCO 上比 VOC 上取得了更高的性能提升. 具体来说,St-CISNet 在 MS COCO 上实现了 3% 的 AP_{50} 改进,而在 PASCAL VOC 上仅获得了 0.6%(1%) 的改进. 这证实了基于位置和时间上下文搜索的强化学习方法可以有效地提高复杂场景中的目标检测能力.

表 3 MS COCO 测试集上的检测结果

单位: %

算法	AP	AP_{50}	AP_{75}	AP_s	AP_M	AP_L	FPS
FasterR-CNN ^[6]	21.9	42.7	—	—	—	—	11
FPN ^[12]	36.2	59.1	39.0	18.2	—	—	1.8
SSD512 ^[8]	27.7	46.4	26.7	10.9	31.8	43.5	19
IR R-CNN ^[17]	39.7	62.0	43.2	21.9	42.4	50.2	—
FA-SSD ^[18]	38.6	—	—	23.3	—	—	18
PANet ^[3]	42.0	65.1	45.7	22.4	44.7	58.1	4.7
RRNet ^[32]	36.4	56.9	37.7	18.8	39.2	51.4	7
St-CISNet*	45.1	67.1	46.8	25.3	47.5	59.3	23

注:*表示多个随机运行的平均结果,加粗数据为最优结果.

对于小目标而言,利用位置和时间上下文指导的强化学习搜索显示出了一定的鲁棒性. 与基于强化学习的目标检测方法相比,St-CISNet 取得了 25.3% 的 AP_s ,高出 RRNet^[31]算法 6%. 与流行的单阶段和两阶段的目标检测方法相比,St-CISNet 将 AP_s 提升了 6%,并且优于加入特征金字塔的 Faster R-CNN. 这得益于嵌入了语义特征和空间特征的双向传递增强了小目标的特征表示,以及自适应多尺度窗口提取方法为小目标的搜索提供了更多的位置上下文信息.

为了全面评估 St-CISNet 的性能,本文在表 3 中给

出了 St-CISNet 的推理速度. 该速度是在装有 NVIDIA GeForce RTX 2080 Ti, CUDA 10.0 和 cuDNN v7 的计算机上进行评估的. 如表 3 所示,St-CISNet 能在 43 ms (23 FPS) 内处理一幅图像. RRNet^[32] 是最新的基于强化学习的目标检测算法,它只能在 142 ms (7 FPS) 内处理一幅图像. 与 RRNet 相比,St-CISNet 速度更快,检测效果更好.

5.4 消融实验

本节在 MS COCO 测试集上进行了“双向特征融合”和“自适应多尺度窗口提取比例”以及“搜索模块

历史信息”的消融实验,结果如表4所示.

表4 MS COCO测试集上的消融实验结果 单位:%

双向特征		自适应多尺度窗口				搜索历史		AP	AP _s
No	Yes	0.5	0.6	0.8	180°	No	Yes		
	√	√					√	37.1	19.2
	√		√				√	37.5	19.6
	√			√			√	37.9	20.1
	√	√	√	√			√	41.5	21.8
√		√	√	√	√		√	40.1	20.2
	√	√	√	√	√	√		40.8	21.4
	√	√	√	√	√		√	45.1	25.3

注:加粗数据为最优结果.

(1)双向特征融合对检测结果的影响.对于双向特征融合,本文在自适应多尺度窗口3种比例、一种变换的条件下进行了两组对比实验,分别为加入双向特征融合和未加入双向特征融合.总体来看,加入特征融合的网络要比未加入的AP_s提高5%,可以看出特征融合是增强目标特征表示的一个重要手段.

(2)自适应多尺度窗口提取比例对检测结果的影响.本文在嵌入双向特征融合的条件下进行了5组实验,前3组分别单独采用0.5,0.6,0.8这3种比例进行了实验,发现结果相差无几,小目标的检测结果也不理想.第4组实验是将3种比例进行了混合,AP和AP_s分别提升了3.4%和2.6%.但是对可视化结果分析发现,MS COCO数据集中原图像与目标的长宽比大都相反.因此,在第5组实验中将3种比例的矩形框进行了180°旋转.结果显示,AP和AP_s都提高了至少4%,这说明对提取框比例的改变可以有效提高小目标的检测精度.

(3)搜索模块输入的历史信息对检测结果的影响.原始动作网络的状态只有当前时刻的观察状态,本文在此基础上加入了多尺度窗口历史信息 H_t 、证据窗口的历史信息 E_t 以及动作的历史信息 G_t .为了验证这些信息的作用,本文对动作网络输入的状态信息进行了两组对比实验,分别是加入历史信息的搜索模块(状态为 f_t)和未加入历史信息的搜索模块(状态为 (H_t, E_t, G_t, f_t)).结果显示,AP和AP_s分别提升了4.3%和3.9%,这说明这些历史信息为搜索目标提供了重要依据.

6 结语

针对小目标检测精度不高的问题,受人眼搜索模式启发,本文提出一种基于位置和时间上下文搜索的强化学习小目标检测方法:St-CISNet.该方法不同于以往的基于强化学习的目标检测方法.St-CISNet通过模拟人眼从整体到局部的搜索方式,使用结合强化学习

的时间上下文搜索模块,利用包含历史信息的时间上下文在含有位置信息的“位置上下文”上逐渐缩小搜索范围,将注意力最终聚焦在小目标上.实验结果表明,St-CISNet比现有方法提高3%,并且在5个搜索次数之内就可以寻找到一个目标的单个实例.后续的工作考虑将背景减除法与模型相结合,以降低St-CISNet对光照变化下小目标检测的敏感度.

参考文献

- [1] MEYE A F, O'KEEFE J, POORT J. Two distinct types of eye-head coupling in freely moving mice[J]. *Current Biology*, 2020, 30(11): 2116-2130.
- [2] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. *Nature*, 2015, 518(7540): 529-533.
- [3] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759-8768.
- [4] LENG J X, REN Y H, JIANG W X, et al. Realize your surroundings: Exploiting context information for small object detection[J]. *Neurocomputing*, 2021, 433: 287-299.
- [5] EVERINGHAM M, VAN GOOL L, WILLIAMS C K I, et al. The pascal visual object classes (VOC) challenge[J]. *International Journal of Computer Vision*, 2010, 88(2): 303-338.
- [6] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common objects in context[C]//European Conference on Computer Vision. Cham: Springer, 2014: 740-755.
- [7] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.
- [8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 779-788.
- [9] LIU W, ANGELOV D, ERHAN D, et al. SSD: Single shot MultiBox detector[C]//European Conference on Computer Vision. Cham: Springer, 2016: 21-37.
- [10] 李宝奇, 贺昱曜, 张伟, 等. 基于并行附加特征提取网络的SSD地面小目标检测模型[J]. *电子学报*, 2020, 48(1): 84-91.

LI B Q, HE Y Y, QIANG W, et al. SSD with parallel ad-

- ditional feature extraction network for ground small target detection[J]. *Acta Electronica Sinica*, 2020, 48(1): 84-91. (in Chinese)
- [11] CAO G M, XIE X M, YANG W Z, et al. Feature-fused SSD: Fast detection for small objects[C]//Proceeding SPIE 10615, Ninth International Conference on Graphic and Image Processing (ICGIP2017). Bellingham: SPIE, 2018: 381-388.
- [12] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 936-944.
- [13] CHEN Z, HUANG S L, TAO D C. Context refinement for object detection[C]//European Conference on Computer Vision. Cham: Springer, 2018: 74-89.
- [14] 冷佳旭, 刘莹. 基于深度学习的小目标检测与识别[J]. *数据与计算发展前沿*, 2020, 2(2): 120-135.
LENG J X, LIU Y. Small object detection and recognition based on deep learning[J]. *Frontiers of Data & Computing*, 2020, 2(2): 120-135. (in Chinese)
- [15] TANG X, DU D K, HE Z Q, et al. PyramidBox: A context-assisted single shot face detector[C]//European Conference on Computer Vision. Cham: Springer, 2018: 812-828.
- [16] PATO L V, NEGRINHO R, AGUIAR P M Q. Seeing without looking: Contextual rescoring of object detections for AP maximization[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2020: 14598-14606.
- [17] FU K, LI J, MA L, et al. Intrinsic relationship reasoning for small object detection[EB/OL]. (2020-09-02) [2022-04-06]. <https://arxiv.org/abs/2009.00833>.
- [18] LIM J S, ASTRID M, YOON H J, et al. Small object detection using context and attention[C]//2021 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). Piscataway: IEEE, 2021: 181-186.
- [19] MNIH V, HEES N, GRAVES A. Recurrent models of visual attention[C]//Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2014: 2204-2212.
- [20] 程旭, 宋晨, 史金钢, 等. 基于深度学习的通用目标检测研究综述[J]. *电子学报*, 2021, 49(7): 1428-1438.
CHENG X, SONG C, SHI J G, et al. A survey of generic object detection methods based on deep learning[J]. *Acta Electronica Sinica*, 2021, 49(7): 1428-1438. (in Chinese)
- [21] KONG T, SUN F C, YAO A B, et al. RON: Reverse connection with objectness prior networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 5244-5252.
- [22] CAICEDO J C, LAZEBNIK S. Active object localization with deep reinforcement learning[C]//2015 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2016: 2488-2496.
- [23] JIE Z Q, LIANG X D, FENG J S, et al. Tree-structured reinforcement learning for sequential object localization [C]//Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2016: 127-135.
- [24] ZHOU M, WANG R J, XIE C J, et al. ReinforceNet: A reinforcement learning embedded object detection framework with region selection network[J]. *Neurocomputing*, 2021, 443: 369-379.
- [25] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2001: I.
- [26] UIJLINGS J R R, VAN DE SANDE K E A, GEVERS T, et al. Selective search for object recognition[J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [27] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE, 2017: 2980-2988.
- [28] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. (2014-09-04)[2022-04-06]. <https://arxiv.org/abs/1409.1556>.
- [29] ITTI L, KOCH C. Computational modelling of visual attention[J]. *Nature Reviews Neuroscience*, 2001, 2(3): 194-203.
- [30] BUENO M B, NIETO X G, MARQUES F, et al. Hierarchical object detection with deep reinforcement learning [J]. *Deep Learning for Image Processing Applications*, 2017, 31(164): 3.
- [31] MATHE S, PIRINEN A, SMINCHISESCU C. Reinforcement learning for visual object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2016: 2894-2902.
- [32] LI Y, HAN X C, GE L T, et al. A recurrent reinforce-

ment learning approach for small object detection with dynamic refinement[C]//2021 International Joint Conference on Neural Networks (IJCNN). Piscataway: IEEE, 2021: 1-8.

作者简介



姜虹 女,1977年出生,陕西宝鸡人.现为西安工业大学计算机科学与工程学院副教授,硕士生导师.主要研究方向为软件工程、图像处理、神经网络与深度学习等.

E-mail: 249479898@qq.com



马姣姣(通讯作者) 女,1997年出生,陕西宝鸡人.现为西安工业大学计算机科学与工程学院硕士研究生.主要研究方向为机器学习、计算机视觉等.

E-mail: 2578516632@qq.com



姚红革 男,1978年出生,陕西西安人.现为西安工业大学计算机科学与工程学院副教授,硕士生导师.主要研究方向为机器学习、计算机视觉、人工智能等.

E-mail: yaohongge@xatu.edu.cn



程嗣怡 男,1980年出生,江苏南京人.现为空军工程大学航空工程学院教授,硕士生导师.主要研究方向为目标检测、电子对抗等.中国电子学会会员编号:E190050619M.

E-mail: csy_316@163.com



陈游 男,1983年出生,湖南岳阳人.现为空军工程大学航空工程学院副教授,硕士生导师.主要研究方向为信息对抗理论与技术等.

E-mail: chenyouisky@163.com



喻钧 女,1971年出生,重庆人.现为西安工业大学计算机科学与工程学院教授,硕士生导师.主要研究方向为图像处理、模式识别等.

E-mail: yujun@xatu.edu.cn