显著性背景感知的多尺度红外行人检测方法

赵 斌 王春平* 付 强

(陆军工程大学石家庄校区电子与光学工程系 石家庄 050003)

摘 要：超大视场(U-FOV)红外成像系统探测范围大、不受光照限制，但存在尺度多样、小目标丰富的特点。为此该文提出一种具备背景感知能力的多尺度红外行人检测方法，在提高小目标检测性能的同时，减少冗余计算。首先，构建了4尺度的特征金字塔网络分别独立预测目标，补充高分辨率细节特征。其次，在特征金字塔结构的横向连接中融入注意力模块，产生显著性特征，抑制不相关区域的特征响应、突出图像局部目标特征。最后，在显著性系数的基础上构建了锚框掩膜生成子网络，约束锚框位置，排除平坦背景，提高处理效率。实验结果表明，显著性生成子网络仅增加5.94%的处理时间，具备轻量特性；超大视场(U-FOV)红外行人数据集上的识别准确率达到了93.20%，比YOLOv3高了26.49%；锚框约束策略能节约处理时间18.05%。重构模型具有轻量性和高准确性，适合于检测超大视场中的多尺度红外目标。

关键词：红外行人检测；超大视场；卷积神经网络；背景感知；多尺度

中图分类号：TN215 **文献标识码**：A **文章编号**：1009-5896(2020)10-2524-09

DOI: 10.11999/JEIT190761

Multi-scale Pedestrian Detection in Infrared Images with Salient Background-awareness

ZHAO Bin WANG Chunping FU Qiang

(*Department of Electronic and Optical Engineering, Shijiazhuang Campus of Army Engineering University, Shijiazhuang* 050003, *China*)

Abstract: The infrared imaging system of Ultrawide Field Of View (U-FOV) has large monitoring range and is not limited by illumination, but there are diverse scales and abundant small objects. For accurately detecting them, a multi-scale infrared pedestrian detection method is proposed with the ability of background-awareness, which can improve the detection performance of small objects and reduce the redundant computation. Firstly, a four scales feature pyramid network is constructed to predict object independently and supplement detail features with higher resolution. Secondly, attention module is integrated into the horizontal connection of feature pyramid structure to generate salient features, suppress feature response of irrelevant areas and enhance the object features. Finally, the anchor mask generation subnetwork is constructed on the basis of salient coefficient to the location of the anchors, to eliminate the flat background, and to improve the processing efficiency. The experimental results show that the salient generation subnetwork only increases the processing time by 5.94%, and has the lightweight characteristic. The Average-Precision is 93.20% on the U-FOV infrared pedestrian dataset, 26.49% higher than that of YOLOv3. Anchor box constraint strategy can save 18.05% of processing time. The proposed method is lightweight and accurate, which is suitable for detecting multi-scale infrared objects in the U-FOV camera.

Key words: Infrared pedestrian detection; Ultrawide Field Of View(U-FOV); Convolutional Neural Network(CNN); Background-awareness; Multi-scale

1 引言

红外成像系统利用物体的热辐射成像，具有抗干扰能力强、不易受恶劣环境影响的优点，能有效弥补可见光设备受光照影响的缺陷，在夜间智能安防、车辆辅助驾驶、军事侦察监视等领域有着广泛的应用前景[1–3]。在此基础上实现的红外行人检测方法充分利用了红外成像系统的优点，具备优异的抗干扰性能，大幅扩展了行人检测技术的应用场景。然而，由于红外成像依赖于物体热辐射，一般

收稿日期：2019-09-30；改回日期：2020-05-13；网络出版：2020-05-20
*通信作者： 王春平 wang_c_p@163.com

存在信噪比低、色彩缺失、纹理细节少的问题[4]，导致红外图像中的行人检测存在较大挑战性。

近年来，学者们提出了大量的行人检测跟踪算法，但多数方法都依赖于人工特征设计，例如纹理[5]、方向梯度直方图[6](Histograms of Oriented Gradients, HOG)、Haar-like[7]以及局部二值模式[8](Local Binary Pattern, LBP)等特征。另一方面，深度卷积神经网(Convolutional Neural Network, CNN)在目标检测领域取得巨大进步，与之相关的行人检测技术也飞速发展。基于CNN的行人检测方法避免了复杂的特征设计环节，可直接由网络自动从数据中学习得到目标的分层特征，无需过多人为参与就能实现端到端的行人检测，但小尺度目标检测精度不佳[9]。Liu等人[10]提出一种阈值自适应的非极大值抑制方法，增强密集行人检测能力。Liu等人[11]将行人检测问题视为高级语义特征检测问题，将任务简化为目标中心与尺度的预测。然而，由于行人尺度的多样性，特别是小尺度行人的大量存在以及红外图像本身的成像缺陷，造成基于深度学习的行人检测方法并没有被广泛应用于红外图像领域，文献[12,13]分别对Fast RCNN[14]和YOLOv3[15]进行改进，实现红外图像行人检测，但并没有解决红外图像中行人特征稀少的问题。

针对红外图像中行人尺度多样、特征有限的问题，本文提出一种具有背景结构感知能力的多尺度行人检测模型。设计了一个4层特征图金字塔目标预测网络，从多个尺度上独立预测目标，补充了更多细节特征，有利于增强小尺度目标的检测能力。然而，多尺度特征图复用[16]会导致计算负担的显著增加，为了抵消高分辨率特征图带来的这一不利影响，本文提出了一种新的锚框(anchor box)生成策略，在选择锚点时，不再逐点遍历特征图，而是通过感知背景排除平坦背景区域，让锚点尽量集中在目标区域附近。为了感知背景，在不同尺度特征图之间通过注意力模块[17,18]构建了目标显著性子网络，由此产生的显著性特征具有抑制背景、突出局部目标特性的能力，在此基础上通过对显著系数进行二值化和局部均值判定就可以生成锚框掩膜，从而达到排除平坦背景区域，减少冗余锚框的目的。

2　行人数据集

实验中主要使用了Caltech行人数据集[19]和超大视场(Ultrawide Field Of View, U-FOV)红外行人数据集。Caltech行人数据集是一个公共数据集，包含约10h、分辨率为640×480、频率为30 Hz的视频，视频由车载摄像机在城市环境中拍摄，总计约250000帧图像，350000标注框和2300个不同的行人。

U-FOV红外行人数据集是一个自建数据集，由于超大视场的成像图像中包含丰富的小尺度目标，使得超大视场红外图像中的多尺度行人检测极具挑战，目前，还未有相关的文献和数据集提出，因此，通过自主采集图像制作了U-FOV红外行人数据集。图像分辨率为800×600，由视场($H\times V$)约为140°×110°，焦距为5.56 mm，成像波段为7～14 μm的红外镜头采集得到。手工标注了1000张训练图像、200张验证图像和661张测试图像。超大视场红外成像系统兼具红外与超大视场相机的优点，能极大改善夜间汽车驾驶视线受阻等问题，有效预防近距离侧方盲区行人突然闯入造成的行车事故。然而在超大视场红外相机捕获的图像中，由于相机的焦距较短，导致目标尺度随距离增加而快速变小，且受限于红外图像对比度低、成像模糊的缺陷，小尺度目标容易淹没在背景中。图1展示了不同距离上的超大视场红外图像行人特性，为了提早感知行人，要求模型对于小尺度目标具有较好的检测性能。

3　目标检测网络结构

基于深度卷积神经网络的目标检测方法虽然取得了巨大进步，但仍存在一些不足：(1)为了获取图像全局信息、增加感受野(receptive field)，网络

(a) 中大尺度行人　　(b) 小尺度行人

图 1 超大视场红外图像行人特性

中存在较大步幅的下采样操作，导致小尺度目标的检测性能受限制；(2)为了提高小尺度目标的检测能力、补充细节特征，常采用多特征层复用或特征金字塔的形式融合多尺度卷积特征，但这势必会造成计算量的大幅增加。同时，发现在Caltech和U-FOV等行人数据集中，都存在较多的道路区域，这些区域的相似性都较高，呈现为平坦背景，属于冗余信息。为了尽量排除这些无用特征，本文提出一种具有平坦背景感知能力的多尺度行人检测网络。

整体网络结构如图2所示，包含主干特征提取网络、显著性生成子网络、锚框掩膜生成子网络和目标预测子网络4部分。主干特征提取网络源自YOLOv3中的Darknet53，用于提取图像的深度卷积特征，为后续目标分类和回归提供特征。为了粗略区分图像的前景和背景区域，本文设计了一个显著性生成子网络，依靠注意力模块对于图像局部特征的敏感性产生显著性特征，该特征具有突出特定任务目标、抑制背景的能力，将其叠加到深度卷积特征中，能在一定程度上弥补小尺度目标特征缺失的问题。针对平坦背景区域显著性系数值较小的特性，构建了一个锚框掩膜生成子网络，在已获取显著性系数的基础上，对其进行二值化及局部领域均值判定处理，从而得到能剔除平坦背景的锚框掩膜，掩膜中为0位置上不产生锚点，可以大大缓解低层高分辨率特征图上预测目标的计算负担。

该结构设计具有以下优点：(1)从4个不同尺度特征图上构建目标预测网络，有效补充了小尺度目标特征信息；(2)将注意力模块搭建在特征金字塔的横向通路上，构建了一个十分轻量的显著性特征提取子网络，对图像局部特征进行增强，有利于提高目标检测性能；(3)利用显著性系数生成的掩膜对特征图进行筛选，排除平坦背景区域，仅在图像的有效区域产生锚框预测目标，可以提高目标检测的执行效率。

3.1 目标预测子网络

原始的YOLOv3网络虽然改善了小目标的检测性能，但仍不足以处理超大视场中的小尺度目标。YOLOv3要求的输入图像分辨率为416×416，用于预测目标的特征图最大分辨率为52×52，这之间存在步长为8的下采样，那么YOLOv3模型理论上能检测到的最小目标分辨率在8×8左右。然而由图1(b)可知，即使在分辨率为800×600的原始图像中都存在小于该理论尺度的目标，因此，需要进一步增大可用于预测目标的特征图分辨率。基于这一考虑，本文重新设计了目标检测网络，增加了一组更低层高分辨率的特征图预测目标，并将其融入到特征金字塔结构中，形成了四尺度的目标预测网络，可以进一步提高小尺度目标的检测精度。训练时，将$y_1 \sim y_4$的预测结果送入损失层中计算损失[15]，指导网络参数调整；检测时，直接在$y_1 \sim y_4$上独立预测目标，得到并整合四个尺度上的检测结果，实现优势互补。

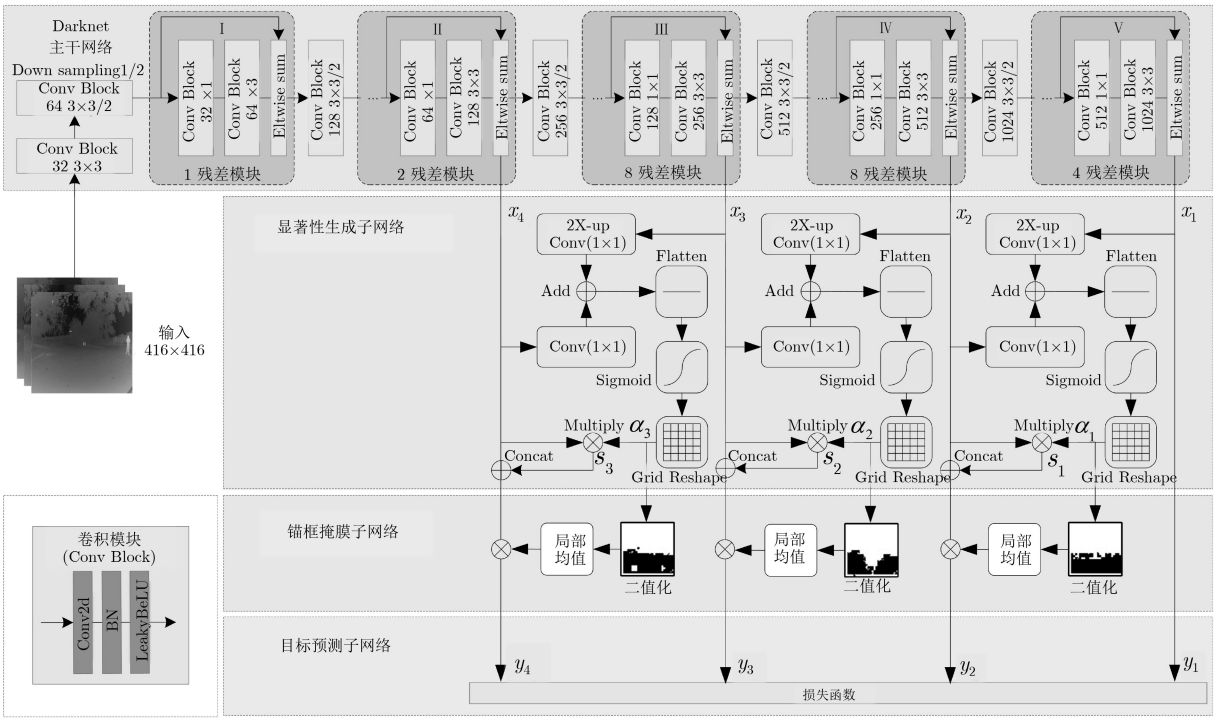


图 2 多尺度红外行人检测网络结构

3.2 显著性生成子网络

显著性生成子网络有两个功能：一是生成显著性特征；二是建立不同尺度特征图之间的横向连接，实现特征融合。显著性特征由注意力模块在两组相邻不同尺度特征图间构造产生，其结构如图3所示。对于两路输入x_1和x_2，首先经过1×1卷积调整成相同通道数，并对低分辨率的特征图x_1进行2倍上采样，将其转换成与x_2通道数相同、分辨率相同的粗略特征图(coarse feature maps)。然后将两路特征图按元素位相加的方式融合，并经激活函数激活后输出。Flatten层将多维输入压平成1维形式，转换成sigmoid需要的输入形式，便于计算显著性系数。最后将介于0～1之间的显著性系数重新网格化为与输入x_2分辨率相同的系数图，并将之与x_2按元素位的方式相乘，生成具有特定局部区域显著特征的特征图。

图4给出了特征图横向连接及特征融合方法，其中α是显著性系数，度量了两组不同尺度特征图之间的相似性，当相似性较高时，即前后不同尺度特征图之间目标继承性较好时，对应区域的显著性系数较大，反之，则显著性系数较小。显著系数图与特征图相乘的过程可视为图像各成分权重重新分配的过程，可以突出重点区域、抑制平坦背景。将生成的显著特征与卷积特征进行拼接，能引导目标预测网络更加关注包含目标的前景区域，有效弥补红外图像特征缺失的问题。该特征融合结构类似于残差结构，输出端：$y=\alpha\cdot x+x$，即使显著性特征即使不起作用，也能保证有普通卷积特征可利用。

3.3 锚框掩膜生成子网络

利用深度卷积网络预测目标时，依赖锚框确定候选目标区域。在已有的检测网络[20,21]结构中，主要采用遍历特征图的方式生成锚点，然而在引入低层高分辨特征图提高小尺度目标的检测性能时，这种遍历的方式会造成计算负担的大幅增加。以YOLOv3为例，它从3个尺度上独立预测目标，每个锚点上生成3个锚框，需要处理的锚框总数量为10647，特征图中一个像素最少对应于输入图像中的8×8图像块，如果要检测出更小的目标，则需进一步增大预测特征图分辨率。但是从4个尺度上预测目标时，则需要额外增加分辨率为104×104的特征图，将产生32448个锚框，这就会引入过多的冗余边框，导致计算量的增加。

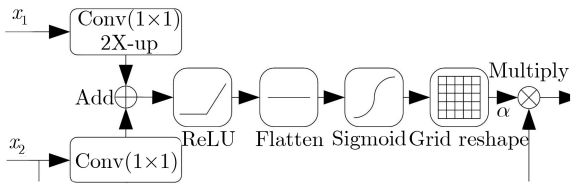


图 3 注意力模块结构

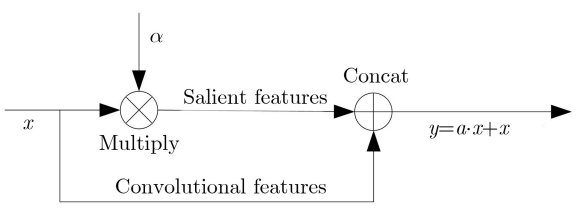


图 4 显著性特征与卷积特征融合方法

针对这一问题，设计了一个锚框掩膜生成子网络，用于排除平坦背景区域，减少冗余锚框生成数量。利用了上一节中的显著性系数α在背景相似性较高区域数值较小的特性，可以在生成锚框时直接排除掉一些不包含目标的平坦背景区域。具体实现如图5所示，锚框掩膜生成子网络通过二值化显著性系数图与局部邻域均值判定产生锚框掩膜。在二值化时，遍历并判定系数图上各像素点的值$\mathrm{img}(i,j)$是否小于设定的阈值T，如式(1)所示，i, j表示像素坐标

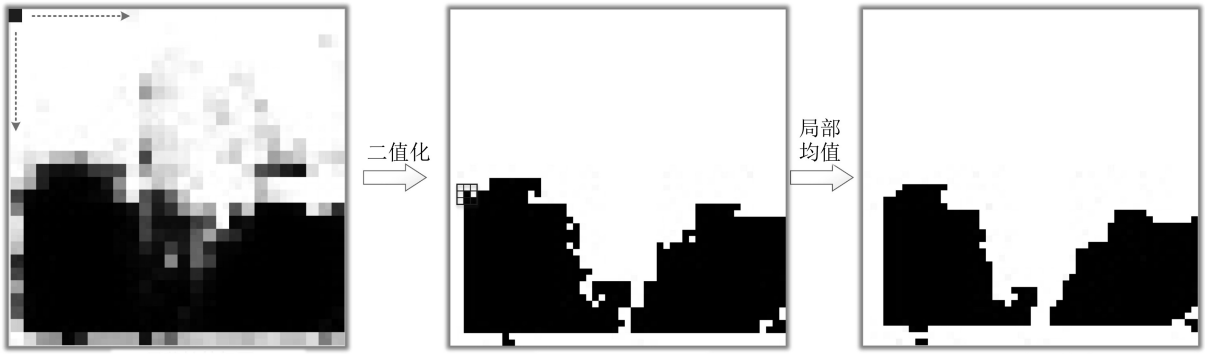
$$\left.\begin{array}{ll}\mathrm{img}(i,j)=0, & \mathrm{img}(i,j)<T\\ \mathrm{img}(i,j)=1, & \mathrm{img}(i,j)\geq T\end{array}\right\} \tag{1}$$

在得到二值化系数图后为了排除某些孤立像素点的影响，计算了0区域内的每个像素点的局部邻域均值，当局部均值大于阈值T_m时，说明该点邻域内0点较少、非0点较多，可能是目标的边缘区域，为了防止影响检测性能，则将该点像素值置为1，否则保持为0，其判定依据如式(2)所示

$$\left.\begin{array}{l}\mathrm{mask\_anchor}(i,j)=0,\\ \quad \mathrm{mean}(\mathrm{img}[i-1:i+1,j-1:j+1])\\ \quad <T_m\&\mathrm{img}(i,j)=0\\ \mathrm{mask\_anchor}(i,j)=1,\ \text{其他}\end{array}\right\} \tag{2}$$

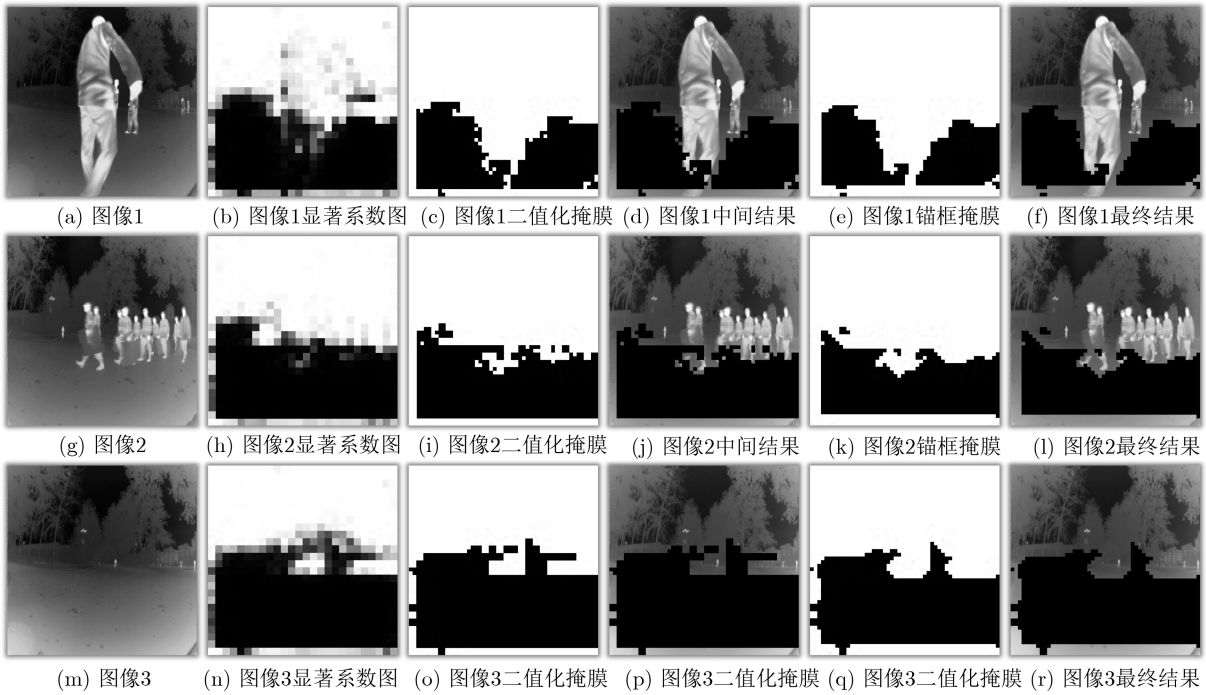
得到锚框掩膜后，将其与特征图进行按位相乘的运算，那么特征图上对应掩膜为0的区域则是背景，在该位置上不产生锚框；对应掩膜为1的区域则是可能存在目标的区域。例如图5(c)中白色区域表示值为1，黑色区域值为0，那么仅在白色区域确定锚点，产生目标候选边框，忽略黑色区域，可以达到缩减冗余边框、提高执行效率的目的，同时保证目标检测性能。

图6是生成锚框掩膜各阶段的处理结果及其对应于原图像中的位置。输入图像经显著性生成子网络处理后生成的显著性系数对应图中的第2列的salient maps。显著性系数图有两个作用：一是直接与深度卷积特征相乘得到显著性特征；二是用于生成锚框掩膜，减少预测边框数量。将介于0～1之间的显著性系数图展开到了0～255之间，二值化阈值



(a) 显著性特征图 (b) 二值化特征图 (c) 锚框掩膜

图 5 锚框掩膜生成过程



(a) 图像1 (b) 图像1显著系数图 (c) 图像1二值化掩膜 (d) 图像1中间结果 (e) 图像1锚框掩膜 (f) 图像1最终结果

(g) 图像2 (h) 图像2显著系数图 (i) 图像2二值化掩膜 (j) 图像2中间结果 (k) 图像2锚框掩膜 (l) 图像2最终结果

(m) 图像3 (n) 图像3显著系数图 (o) 图像3二值化掩膜 (p) 图像3二值化掩膜 (q) 图像3二值化掩膜 (r) 图像3最终结果

图 6 不同输入图像的锚框掩膜

设为$T=100/255\approx0.39$；局部均值判定阈值为$T_m=2/9$，局部均值判定阈值的大小约束了局部邻域非0点的数量，此阈值设定下容许局部8邻域内最多存在两个非0点。

对比图6最后一列结果可知，生成的锚框掩膜对于不同尺度的红外行人目标都具有很好的敏感性，能有效抑制图像中的平坦背景且不会对前景目标造成影响。第4列和第6列结果是由对应的二值化掩膜和锚框掩膜按比例放大后与输入图像按位相乘得到的，横向对比这两列可知，局部均值处理过程能约束掩膜边缘，排除局部孤立区域的干扰。注意到掩膜对于小目标保留较好，但可能会覆盖一些大尺度行人目标的局部边缘，这对检测性能造成影响却不大，有两方面原因：一是大尺度目标的检测性能主要依赖于最后一层的卷积特征，这一层特征图分辨率小，不需要对其锚点进行削减，因此不存在掩膜干扰的问题；二是大尺度目标特征丰富，即使损失部分目标边缘细节，也能保证目标的检测性能。

4 实验结果与分析

4.1 实验设置

实验环境为64位Windows操作系统，NVIDIA GeForce GTX TITAN X GPU；软件采用Keras，并以Tensorflow为后端进行卷积神经网络计算；编程语言为Python3.6。以YOLOv3模型为基本框架重新设计检测网络，在经ImageNet和COCO数据集预训练的Darknet53上构建行人检测模型，使用Adam优化算法进行训练。训练中采用了两次迁移训练的方式进一步提高性能：第1次迁移训练在Caltech行人检测数据集上进行，旨在扩充数据量和增强模型对于行人类目标的鲁棒性与泛化能力；

第2次迁移训练在 U-FOV红外行人数据集上进行，训练模型对于红外图像目标的识别定位能力。

4.2 显著性系数图及锚框掩膜

为了检验模型对于行人目标的敏感性和对于不同数据集的泛化能力，在图7中给出了不同二值化阈值条件下的锚框掩膜生成情况。其中，图7(a)的输入图像是一张合成图像，在图6(m)输入图像的平坦背景中增加了3个小尺度目标，其目的是检验显著性生成子网络能否正确感知平坦背景中包含的目标，结果表明模型能有效验证模型对于行人目标、特别是小尺度行人目标的感知能力。

对比图6(r)与图7(f)可知，显著性生成子网络能根据图像内容调整显著性系数，体现了注意力模块对平坦背景的抑制能力和对图像局部目标特征的增强作用，也进一步证明了建立在显著性系数基础上的锚框掩膜生成网络的可行性。图7(g)和图7(m)的输入图像分别来自LTIR(Linköping Thermal InfraRed)数据集[22]和Caltech行人数据集(可见光)，对应生成的锚框掩膜都具有抑制平坦背景的特性，说明模型具有较好的泛化能力，而不仅仅在特定环境数据集上有效。第3～5列给出了不同二值化阈值条件下的锚框掩膜生成结果，当阈值较小时，几乎不会损失目标信息，但对背景的抑制范围相对较少；而当阈值过大时，对背景抑制较多，但可能会覆盖更多的目标边缘，对检测性能造成影响。

4.3 超大视场红外行人检测性能评估

图8是U-FOV红外行人数据集中5幅包含不同尺度行人目标的测试图像可视化检测结果。对比不同方法的处理结果可知，本文所提方法取得了更好的小尺度目标检测效果，检测结果的最小边框大小为6×13，测试图像图8(a)的结果验证了这一结论，其他方法都不能完全检测出该图中的5个小尺度行人。对比图8(b)—图8(e)结果可知，SSD[23]由于采用Mobilenet_v1作为特征提取主干网络，虽然具有较快的处理速度，但其综合检测性能最差，说明深度卷积特征的好坏对检测结果至关重要；Faster R-CNN[24]与R-FCN[25]没有复用多尺度特征图，仅在最后一个尺度特征图上预测目标，使得检测网络缺乏细粒度，容易将多个密集目标回归到一个目标边框中，造成目标的漏检；YOLOv3[13]虽然从3个尺度上预测目标，但仍然不能解决尺度过小目标的检测问题；CSP[11]直接通过卷积操作预测行人的中心位置和尺度大小，虽然在检测率和处理速度上有大幅提升，但其目标定位精度有所下降，且在红外小尺度目标检测上仍存在较多漏检；本文所提方法则在从大到小多个尺度的目标检测问题上表现优异，适合用于处理超大视场图像数据中的行人目标。

为了定量评价不同方法在U-FOV数据集上的表现，对测试集中的661幅测试图像进行检测，其结果如表1所示。表中“+FS”表示增加四尺度目标预测子网络；“U-FOV”表示仅在U-FOV数据集上做迁移训练；“Caltech+U-FOV”表示先在Caltech数据集上做第1次迁移训练，再在U-FOV数据集上做第2次迁移训练。评价指标采用平均准确

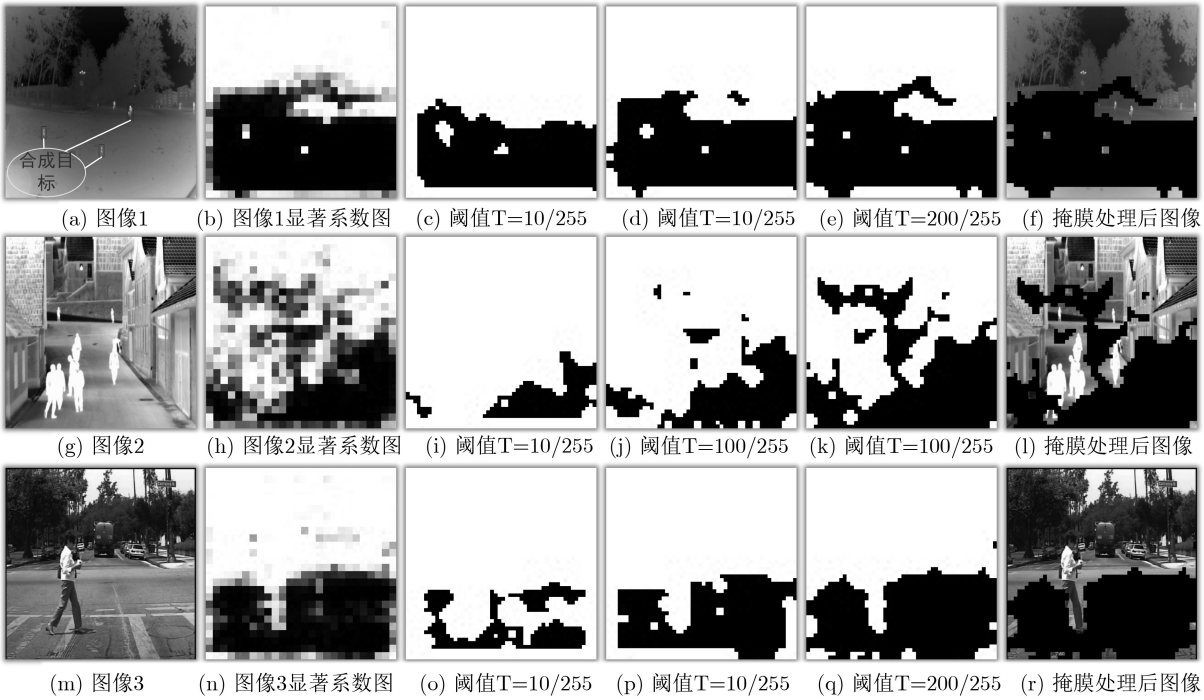
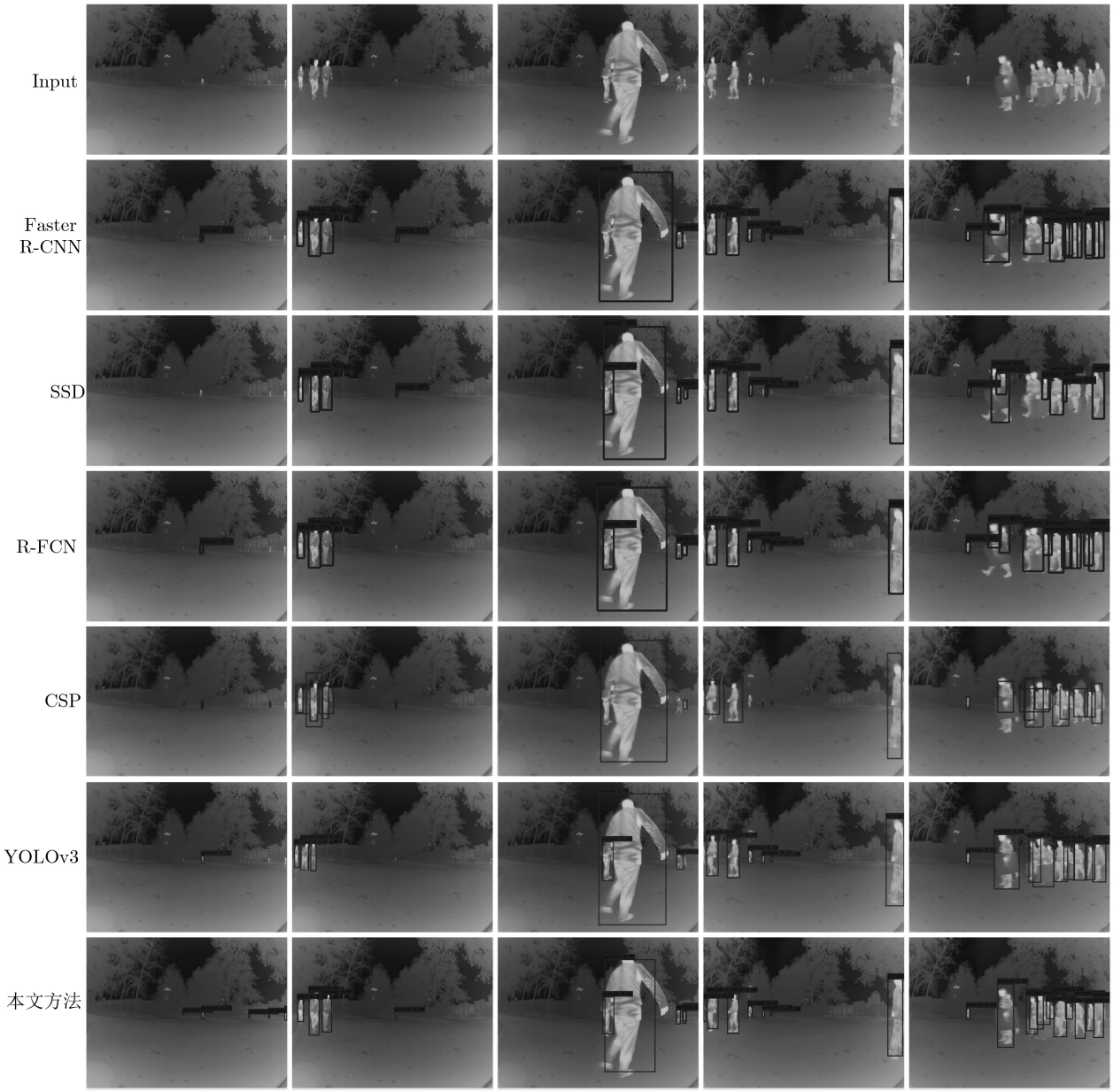


图 7 不同二值化阈值下的锚框掩膜



(a) 测试图像1及检测结果 (b) 测试图像2及检测结果 (c) 测试图像2及检测结果(d) 测试图像2及检测结果(e) 测试图像5及检测结果

图 8 红外行人检测可视化结果

表 1 不同IoU阈值下的行人检测平均准确率

方法	主干网络	训练集	平均准确率(AP)			
			IoU=0.3	IoU=0.45	IoU=0.5	IoU=0.7
Faster R-CNN	ResNet101	U-FOV	–	–	0.5932	–
SSD	Mobilenet_v1	U-FOV	–	–	0.5584	–
R-FCN	ResNet101	U-FOV	–	–	0.6312	–
CSP	Resnet50	U-FOV	–	–	0.8414	–
YOLOv3	Darknet53	U-FOV	0.6595	0.6671	0.6628	0.6461
YOLOv3+FS	Darknet53	U-FOV	0.8880	0.8870	0.8828	0.8511
YOLOv3+FS	Darknet53	Caltech+U-FOV	0.9057	0.9078	0.9084	0.8961
本文方法	Darknet53	Caltech+U-FOV	**0.9201**	**0.9320**	**0.9315**	**0.9107**

率(Average-Precision, AP)，衡量了模型对于行人类别的检测能力，由precision-recall曲线计算得到。通过设置不同的IoU(Intersection over Union)阈值得到了不同的AP值，并保留了置信度分数大于0.3的检测结果。前3种方法都将IoU阈值设为0.5，因此，仅统计了该条件下的AP值。

由于Faster R-CNN和R-FCN仅在最后一层特征图上预测目标，导致小目标检测性能较低；SSD采用的主干网络泛化性能不够好，使得红外行人的不变特征不容易被完全提取；CSP的目标检测性能需要依靠丰富的行人特征来生成响应图，然而红外图像的特征丰富性远低于可见光图像，导致该方法的检测性能和定位精度都收到影响；YOLOv3虽然已经利用了特征金字塔结构，但特征图的分辨率仍不足以满足微小目标检测的需求；通过引入更低层高分辨率的特征图建立四尺度目标预测网络结构，大幅提高了在U-FOV数据集上的检测性能，说明低层特征对于微小目标检测至关重要。当IoU为0.45时，此时，所提方法在U-FOV红外行人测试集上的平均准确率达到了93.20%，比YOLOv3高26.49%，这些增益中，四尺度预测网络贡献了21.99%，两次迁移训练方式贡献了2.08%，显著性生成子网络和锚框掩膜子网络联合贡献了2.67%。在加入低层高分辨率的特征图做四尺度的目标预测后，检测性能得到明显改善，主要原因在于U-FOV数据集存在大量尺度过小的行人目标，导致YOLOv3原始模型存在过多漏检。

表2统计了所提方法与YOLOv3模型的参数量对比情况，在增加注意力模块和掩膜锚框生成模块后，网络总参数量仅增加了约5.34%，证明了子网络的轻量特性。同时在表3中给出了遍历一次U-FOV测试集661张测试图像的总耗时及处理帧速，由于每次测试时间存在微小偏差，因此取了10次测试总时间的平均值。“+Attention”表示在模型上增加注意力模块构建显著性子网络，目的在于测试注意力模块带来的计算负担，处理时间增加了约5.94%，基本与网络参数的增加量相对应。“FS+Attention”则表示不包含锚框掩膜子网络的四尺度目标预测模型，由于增加了更高分辨率特征图预测目标，处理时间增加了约30.10%。本文所提方法在加入锚框掩膜后，能节约处理时间约18.05%，且能保证多尺度红外目标的高检测性能。

表 2　参数量对比

方法	总参数量	可训练参数量	不可训练参数量
YOLOv3	61576342	61523734	52608
本文方法	64861976	64806296	55680

表 3　U-FOV测试集图像总处理时间及处理帧速

方法	YOLOv3	YOLOv3+Attention	FS+Attention	本文方法
总时间(s)	90.35	95.72	125.39	107.25
处理帧率	7.32	6.91	5.27	6.16

5　结束语

针对超大视场系统中红外行人的成像特点，本文提出一种基于显著性特征的背景感知方法，在提高小尺度目标的检测性能的同时，防止引入过多计算负担。在构建多尺度目标检测框架时，从4个尺度特征图上独立预测目标，增加了检测网络的细粒度；同时，设计了显著性和锚框掩膜两个子网络，将其融入到特征金字塔的横向连接中，保证了子网络的轻量性，并能抑制平坦背景、突出重点目标。然而，主干特征提取网络的卷积计算仍旧占据较多计算开销，需要进一步精简主干网络结构、减少冗余计算，红外行人检测网络的处理效率仍有较大的提升空间。

参 考 文 献

[1] BLOISI D D, PREVITALI F, PENNISI A, *et al*. Enhancing automatic maritime surveillance systems with visual information[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(4): 824–833. doi: 10.1109/TITS.2016.2591321.

[2] KANG J K, HONG H G, and PARK K R. Pedestrian detection based on adaptive selection of visible light or far-infrared light camera image by fuzzy inference system and convolutional neural network-based verification[J]. *Sensors*, 2017, 17(7): 1598. doi: 10.3390/s17071598.

[3] KIM S, SONG W J, and KIM S H. Infrared variation optimized deep convolutional neural network for robust automatic ground target recognition[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 195–202. doi: 10.1109/CVPRW.2017.30.

[4] 王晨, 汤心溢, 高思莉. 基于人眼视觉的红外图像增强算法研究[J]. 激光与红外, 2017, 47(1): 114–118. doi: 10.3969/j.issn.1001-5078.2017.01.022.

WANG Chen, TANG Xinyi, and GAO Sili. Infrared image

enhancement algorithm based on human vision[J]. *Laser & Infrared*, 2017, 47(1): 114–118. doi: 10.3969/j.issn.1001-5078.2017.01.022.

[5] MUNDER S, SCHNORR C, and GAVRILA D M. Pedestrian detection and tracking using a mixture of view-based shape-texture models[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2008, 9(2): 333–343. doi: 10.1109/TITS.2008.922943.

[6] DALAL N and TRIGGS B. Histograms of oriented gradients for human detection[C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, USA, 2005: 886–893. doi: 10.1109/CVPR.2005.177.

[7] ZHANG Shanshan, BAUCKHAGE C, and CREMERS A B. Informed haar-like features improve pedestrian detection[C]. 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 947–954. doi: 10.1109/CVPR.2014.126.

[8] WATANABE T and ITO S. Two co-occurrence histogram features using gradient orientations and local binary patterns for pedestrian detection[C]. The 2nd IAPR Asian Conference on Pattern Recognition, Naha, Japan, 2013: 415–419. doi: 10.1109/ACPR.2013.117.

[9] 余春艳, 徐小丹, 钟诗俊. 面向显著性目标检测的SSD改进模型[J]. 电子与信息学报, 2018, 40(11): 2554–2561. doi: 10.11999/JEIT180118.
YU Chunyan, XU Xiaodan, and ZHONG Shijun. An improved SSD model for saliency object detection[J]. *Journal of Electronics & Information Technology*, 2018, 40(11): 2554–2561. doi: 10.11999/JEIT180118.

[10] LIU Songtao, HUANG Di, and WANG Yunhong. Adaptive NMS: Refining pedestrian detection in a crowd[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 6452–6461. doi: 10.1109/CVPR.2019.00662.

[11] LIU Wei, LIAO Shengcai, REN Weiqiang, *et al*. Center and scale prediction: A box-free approach for pedestrian and face detection[C]. The IEEE Conference on Computer Vision and Pattern Recognition, Los Angeles, USA, 2019: 5187–5196.

[12] 车凯, 向郑涛, 陈宇峰, 等. 基于改进Fast R-CNN的红外图像行人检测研究[J]. 红外技术, 2018, 40(6): 578–584. doi: 10.11846/j.issn.1001_8891.201806010.
CHE Kai, XIANG Zhengtao, CHEN Yufeng, *et al*. Research on infrared image pedestrian detection based on improved fast R-CNN[J]. *Infrared Technology*, 2018, 40(6): 578–584. doi: 10.11846/j.issn.1001_8891.201806010.

[13] 王殿伟, 何衍辉, 李大湘, 等. 改进的YOLOv3红外视频图像行人检测算法[J]. 西安邮电大学学报, 2018, 23(4): 48–52. doi: 10.13682/j.issn.2095-6533.2018.04.008.
WANG Dianwei, HE Yanhui, LI Daxiang, *et al*. An improved infrared video image pedestrian detection algorithm[J]. *Journal of Xi'an University of Posts and Telecommunications*, 2018, 23(4): 48–52. doi: 10.13682/j.issn.2095-6533.2018.04.008.

[14] GIRSHICK R. Fast R-CNN[C]. 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1440–1448. doi: 10.1109/ICCV.2015.169.

[15] REDMON J and FARHADI A. YOLOv3: An incremental improvement[EB/OL]. http://arxiv.org/abs/1804.02767, 2018.

[16] 郭智, 宋萍, 张义, 等. 基于深度卷积神经网络的遥感图像飞机目标检测方法[J]. 电子与信息学报, 2018, 40(11): 2684–2690. doi: 10.11999/JEIT180117.
GUO Zhi, SONG Ping, ZHANG Yi, *et al*. Aircraft detection method based on deep convolutional neural network for remote sensing images[J]. *Journal of Electronics & Information Technology*, 2018, 40(11): 2684–2690. doi: 10.11999/JEIT180117.

[17] CHEN Long, ZHANG Hanwang, XIAO Jun, *et al*. SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 6298–6306. doi: 10.1109/CVPR.2017.667.

[18] WOO S, PARK J, LEE J Y, *et al*. CBAM: Convolutional block attention module[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: 10.1007/978-3-030-01234-2_1.

[19] DOLLÁR P, WOJEK C, SCHIELE B, *et al*. Pedestrian detection: An evaluation of the state of the art[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012, 34(4): 743–761. doi: 10.1109/TPAMI.2011.155.

[20] FU Chengyang, LIU Wei, RANGA A, *et al*. DSSD: Deconvolutional single shot detector[J]. arXiv, 2017, 1701.06659.

[21] HE Kaiming, GKIOXARI G, DOLLÁR P, *et al*. Mask R-CNN[C]. 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 2980–2988. doi: 10.1109/ICCV.2017.322.

[22] BERG A, AHLBERG J, and FELSBERG M. A thermal object tracking benchmark[C]. The 2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance, Karlsruhe, Germany, 2015: 1–6. doi: 10.1109/AVSS.2015.7301772.

[23] LIU Wei, ANGUELOV D, ERHAN D, *et al*. SSD: Single shot multibox detector[C]. Proceedings of the 14th European Conference on Computer Vision, Amsterdam, Netherlands, 2016: 21–37. doi: 10.1007/978-3-319-46448-0_2.

[24] REN Shaoqing, HE Kaiming, GIRSHICK R, *et al*. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137–1149. doi: 10.1109/TPAMI.2016.2577031.

[25] DAI Jifeng, LI Yi, HE Kaiming, *et al*. R-FCN: Object detection via region-based fully convolutional networks[C]. Advances in Neural Information Processing Systems, Barcelona, Spain, 2016: 379–387.

赵 斌：男，1990年生，博士生，研究方向为深度学习、目标检测.
王春平：男，1965年生，博士生导师，研究方向为图像处理、火力控制理论与应用.
付 强：男，1981年生，讲师，博士，研究方向为计算机视觉、网络化火控与指控技术.

责任编辑：余 蓉