引用格式：ZHANG Penghui, LIU Zhi, ZHENG Jianyong, et al. Real-time Infrared Target Detection Algorithm for Embedded System in Complex Scene[J]. Acta Photonica Sinica, 2022, 51(2):0210002
张鹏辉,刘志,郑建勇,等.面向嵌入式系统的复杂场景红外目标实时检测算法[J].光子学报,2022,51(2):0210002

面向嵌入式系统的复杂场景红外目标实时检测算法

张鹏辉[1],刘志[2],郑建勇[3],何博侠[1],裴雨浩[1]
(1 南京理工大学 机械工程学院,南京 210094)
(2 南京博蓝奇智能科技有限公司,南京 210014)
(3 上海大学 人工智能研究院,上海 200444)

摘　要：为了解决复杂背景条件下,红外目标检测存在的准确率低、召回率低、以及网络模型在嵌入式计算平台上推理速度慢的问题,以轻量化网络YOLOv4-Tiny作为算法的基本架构,结合视觉注意力机制和空间金字塔池化思想,提出两种面向嵌入式系统的红外目标检测网络,利用迁移学习策略进行训练,在以昇腾310 AI芯片为核心的Atlas 200 DK嵌入式计算平台进行部署。实验结果表明,在该嵌入式计算平台上推理分辨率为640 pixel×512 pixel的红外图像,相较于原始网络YOLOv4-Tiny,所提网络YOLOv4-Tiny+SE+SPP的平均准确率和召回率分别提升12.36%和18.6%,推理速度达到78 fps;所提网络YOLOv4-Tiny+CBAM+SPP的平均准确率和召回率分别提升15.94%和22.89%,推理速度达到71 fps,可兼顾准确率和实时性,能够满足军事和安防领域对红外目标进行实时检测和跟踪的需要。

关键词：红外图像;注意力机制;迁移学习;目标检测;嵌入式平台
中图分类号：TP391.4　　**文献标识码**：A　　**doi**：10.3788/gzxb20225102.0210002

0 引言

红外辐射穿透力强,不易被云雾吸收,也不受天气干扰,在阴雨天和夜晚都能正常成像,因而,红外视觉在军事和安防领域有着独特的优势,但是,相较于可见光图像,红外图像有着对比度低,分辨率低,目标细节不够丰富的缺点。当前,针对可见光图像的深度学习目标检测算法已取得丰硕的成果[1-3],虽然这些算法也可用于红外目标的检测,但需要解决红外图像对比度低、目标细节模糊带来的准确率降低的问题,尤其当目标成像背景比较复杂时,需要设法提高算法的抗干扰能力,解决召回率降低的问题。另一方面,深度学习技术在民用领域可选择的部署平台较多,而在军事和安防领域,很多情况下只能部署于嵌入式平台,且对算法推理的实时性有很高的要求。因此,针对复杂场景中的红外目标,研究面向嵌入式系统的实时检测算法很有必要。

在基于深度学习的目标检测算法中,一阶段法能够在检测速度和检测精度之间取得较好的平衡,代表性算法有YOLO[4],SSD[5],RetinaNet[6]等,其中YOLO系列算法以其推理速度快的优势,常被用来部署于嵌入式系统。文献[7]提出一种基于YOLOv3架构的汽车目标检测模型,部署于NVIDIA Jetson TX1嵌入式计算平台上,在线检测速度可达到23 fps。文献[8]基于深度学习技术提出一种面向嵌入式设备的快速人群计数算法,设计了弱算力平台加速网络,对分辨率为640×480的图像,在cortex-A72双核ARM平台上可获得20 fps的推理速度。总之,已有研究虽然实现了深度神经网络在嵌入式系统上的部署应用,但总体来看,

基金项目：国家自然科学基金(No. 51575281),中央高校基本科研业务费专项资金(No. 30916011304)
第一作者：张鹏辉(1996—),男,硕士研究生,主要研究方向为机器视觉、模式识别和深度学习。Email: xhh0608@foxmail.com
导师(通讯作者)：何博侠(1972—),男,副教授,博士,主要研究方向为机器视觉、工业人工智能。Email: heboxia@163.com
收稿日期：2021-08-14;**录用日期**：2021-11-04

其推理速度依然比较低，还不能满足军事和安防领域对动态目标进行实时检测和跟踪的需要，因此，研究面向嵌入式计算平台，能够兼顾实时性和准确率的网络模型，具有重要的应用价值。

红外图像不像可见光图像有较高的对比度和丰富的目标细节特征，特别是当背景比较复杂，或者存在部分遮挡时，极易引起目标识别率和召回率降低[9]。文献[10]针对复杂背景和低信噪比条件下的红外弱小无人机目标检测场景，改进了一种含空间注意力机制的红外弱小目标检测网络，提升了网络准确率。文献[11]提出了一种改进的YOLOv3红外末制导目标检测方法，采用预训练思想和Adam梯度下降算法，提高了模型的平均准确率(Mean Average Precision，mAP)，虚警率和漏检率都得到降低，但在PC平台上的检测速度仅有25 fps。文献[12]基于YOLOv3网络，提出了一种深度注意力机制的多尺度红外行人检测网络，并结合迁移学习方法进行网络训练，平均准确率比YOLOv3提高26.74%，适用于多尺度红外行人检测场景。这些研究提升了网络的准确率，但是推理速度比较慢，同样不能满足军事和安防领域的实时性需要。

本文研究面向嵌入式系统的复杂场景红外目标实时检测算法，旨在达到检测速度和平均准确率的良好平衡。选择轻量化网络YOLOv4-Tiny作为检测框架，利用视觉注意力机制对有效通道进行选择，加强对目标的关注度；采用空间金字塔池化结构对注意力机制加强后的特征层进行不同尺度的特征融合，丰富特征图的表达能力；应用迁移学习方法，进一步提高模型的准确率和召回率；最后在嵌入式计算平台进行网络模型部署，验证所提方法的有效性。

1 方法原理

1.1 YOLOv4-Tiny算法架构

YOLOv4-Tiny是YOLOv4[13]的轻量化版本，其网络结构如图1所示。主干特征提取网络CSPDarkNet53_Tiny由2层CBL和3层CSP组成，其中，CBL层由卷积层Conv2d、归一化处理层BatchNorm2d、激活层LeakyReLU这3部分组成，CSP层由4个CBL层和一个最大池化MaxPool2d组成。特征提取网络末端使用两个特征层进行分类与位置回归预测，通过特征金字塔进行特征融合，输出两个检测头Yolo Head。

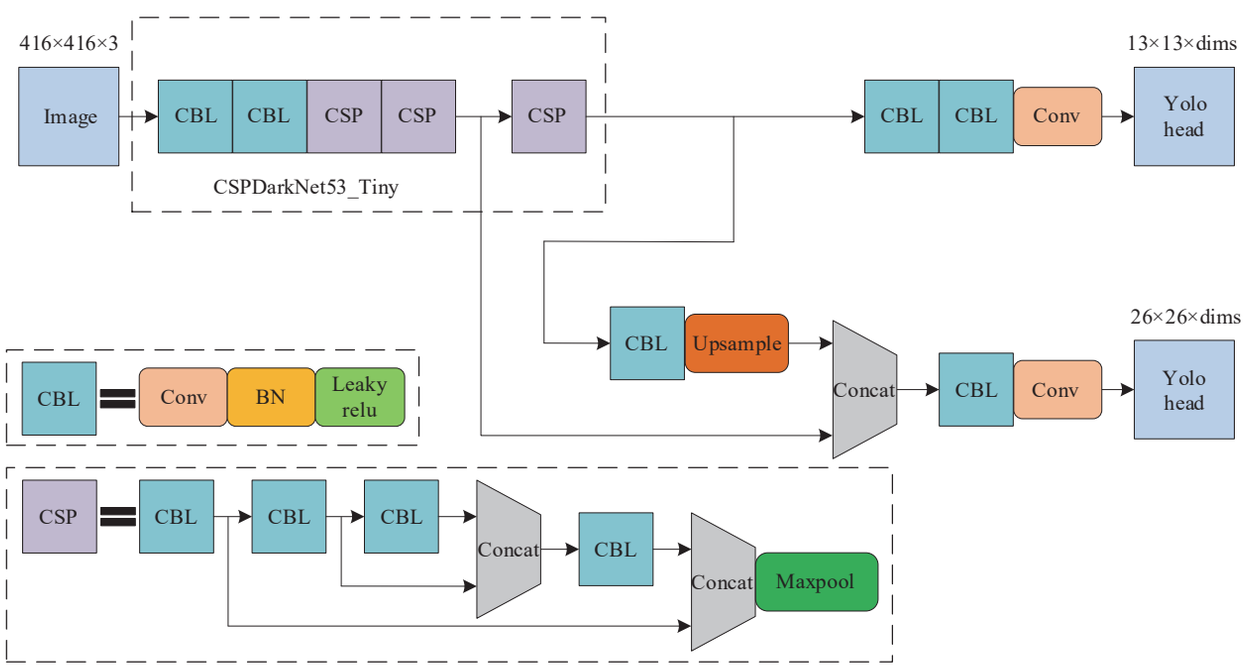

图1 YOLOv4-Tiny网络结构

Fig.1 The network structure of YOLOv4-Tiny

1.2 红外目标检测网络

本文所提红外目标检测网络结构如图2所示。红外复杂场景的目标检测中存在着大量的背景干扰信

息，为了加强对目标的关注度，降低背景无关信息的影响，利用视觉注意力机制有效学习到特征图的权重分布，对特征图进行重新标定，加强对目标的关注度，提高模型的检测识别能力。SE[14](Squeeze-and-Excitation，SE)模块，可以学习到通道之间的依赖关系，在通道维度对原始特征进行重新标定，增强网络对有效通道特征的敏感度；CBAM[15](Convolutional Block Attention Module，CBAM)，利用特征图的通道和空间信息，在这两个维度计算原始特征图的注意力权重图，然后将注意力权重图赋予原始特征图，达到特征自适应学习的目的。由于红外目标轮廓模糊、图像对比度低，为了加强多尺度特征融合，采用SPP[16](Spatial Pyramid Pooling，SPP)模块，通过窗口大小分别为5、9、13对特征图进行池化，将多尺度特征进行融合，丰富特征图的信息，提高不同尺度红外目标的识别和定位能力。

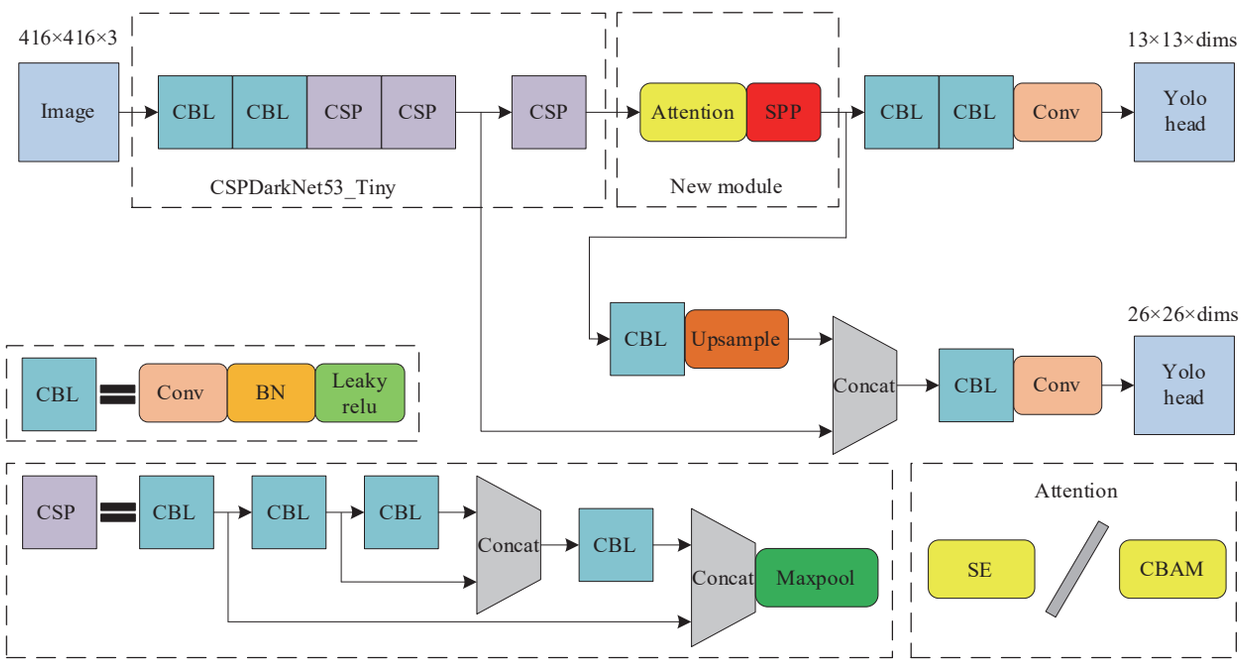

图 2　YOLOv4-Tiny＋SE＋SPP/YOLOv4-Tiny＋CBAM＋SPP 网络结构

Fig.2　The network structure of YOLOv4-Tiny＋SE＋SPP/YOLOv4-Tiny＋CBAM＋SPP

本文选取了在复杂背景下密集人群的样例图像，利用Grad-CAM[17]对经过注意力机制加强后的特征图Yolo Head(13×13×dims)进行了可视化，如图3所示，采用热力图的形式进行可视化表示，温度越高的地方代表网络越关注的地方，反之代表网络不关注的地方。在图3(b)中是未加入注意力机制的原始网络，可以发现其关注点比较分散，包含了较多的无关复杂背景信息；在图3(c)和(d)中加入注意力机制后，网络可以有效地区分于周围的复杂背景干扰信息，将注意力集中于目标中。可视化结果表明，通过加入注意力机制有效地提高对红外目标的关注度。

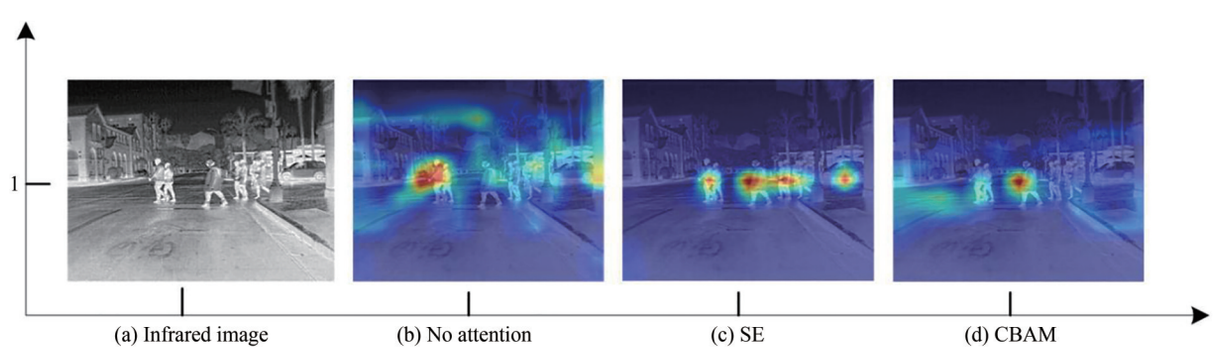

图 3　Grad-CAM 网络可视化结果

Fig.3　Grad-CAM network visualization results

2 实验与结果分析

算法的评价指标采用标准的PASCAL VOC，即平均准确率mAP，用于实验对比和分析。

2.1 数据集

实验采用FLIR公司发布的公开红外数据集[18]。该数据集是采用红外热像仪在天气为晴到多云的加利福尼亚州圣巴巴拉市街道和公路上，时间为11月至次年5月期间的日间(60%)和夜间(40%)进行采集的。红外分辨率为640×512，标注类别为人、汽车、自行车、狗和其他，由于狗和其他类别的数量占比非常少，为了避免由于类别数量的严重不平衡对算法评估产生不必要的影响，因此对数据进行了预处理，去除了狗和其他类别，最终得到用于算法评估的3类目标的红外数据集，其中，训练集和验证集为7 859张图片，测试集为1 360张图片，各类别数量分布如图4所示。

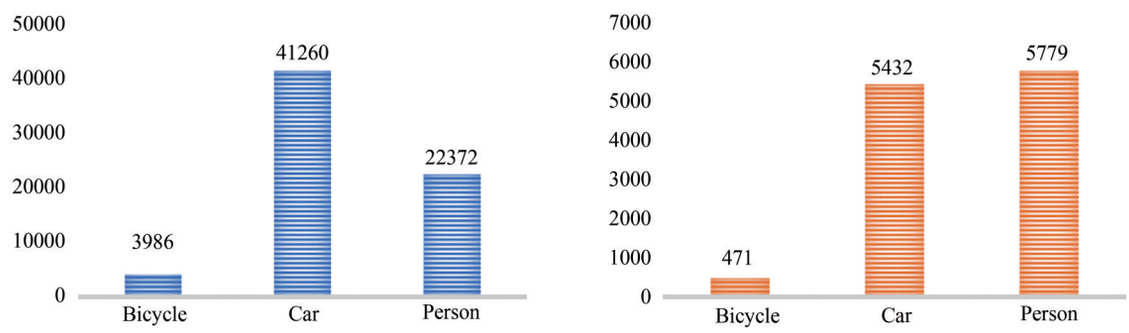

图 4 训练集与验证集(左)和测试集(右)中不同类别数量分布情况

Fig.4 Quantity distribution of different categories in training and validation sets(left)test set(right)

由于FLIR红外数据集是基于车载光电设备，在公路行驶场景下，昼夜进行采集的，因此拥有着复杂背景、部分遮挡、密集目标、不同尺度的特性，部分红外和可见光图像如图5所示。四种场景会相互包含多种特性，以其最突出的特性作为场景特性，场景(a)：夜间条件，由于复杂背景的影响，易造成目标漏检问题；场景(b)：白天条件，背景与目标温度差异较小，并存在目标之间相互遮挡，极易产生识别定位不准确、漏检问题；场景(c)：白天条件，大量密集重叠的目标，产生较多的漏识别问题；场景(d)：夜间条件，场景中存在着尺度不同的目标，出现目标识别定位不准确问题。因此，需要在该红外数据集上进行算法的改进，以解决定位不准确、漏识别问题。

图 5 四种不同场景下的红外和可见光图像

Fig.5 Infrared and visible images of four different scenes

2.2 实验

算法模型采用深度学习框架PyTorch，利用Python语言进行实现。实验训练平台采用台式计算机，系统为Ubuntu18.04.5 LTS，CPU为i7-9700K，GPU为GeForce RTX 2080Ti，内存大小为32 GB，CUDA版本为10.2。

为了保证实验结果的有效性，采用的训练超参数均保持一致。一批图片的数量设置为64张，初始化学习率设置为0.001，使用Adam梯度下降算法，学习率优化策略采用余弦退火算法。数据增强除了采用传统的图像随机翻转、裁切、缩放、色域变换、加入随机噪声外，使用了马赛克数据增强方法。

2.2.1 注意力机制对比实验

对原始网络YOLOv4-Tiny和加入不同的注意力机制模块后重新设计的网络YOLOv4-Tiny+SE和YOLOv4-Tiny+CBAM，在训练集和验证集上进行训练120轮后，在测试集上进行测试，注意力机制对比实验结果如表1所示。

表1 注意力机制对比实验结果

Table 1 Results of attentional mechanism ablation experiment

Model	AP/%			Precision/%	Recall/%	mAP/%	FPS
	Bicycle	Car	Person				
YOLOv4-Tiny(original)	32.85	78.65	61.84	**82.69**	41.33	57.78	**232**
YOLOv4-Tiny+SE	38.47	**80.79**	66.61	80.35	50.18	61.95	226
YOLOv4-Tiny+CBAM	**38.97**	80.34	**66.70**	79.65	**50.48**	**62.00**	215

由表1结果中可以看出，增加了注意力机制，由于整个网络的训练参数量略微增加导致模型推理速度有所下降，相较于YOLOv4-Tiny，增加SE模块后检测速度从232 fps下降至226 fps，减少6 fps；增加CBAM模块后检测速度从226 fps下降至215 fps，减少17 fps。在检测性能方面，加入了两种注意力机制后，召回率分别提升8.85%和9.15%，较大地改善了自行车和人两类的AP；但准确率有所下降，分别下降2.34%和3.04%。实验结果表明，视觉注意力机制能够对有效通道进行选择，应用在FLIR红外数据集中，有效地提升召回率和平均准确率。

2.2.2 空间金字塔池化对比实验

上节实验结果表明，设计的视觉注意力机制网络可以有效地提升mAP。本节中，通过加入空间金字塔池化结构SPP模块对注意力机制增强后的特征图进行全局特征和局部特征的融合，扩大了特征图的感受野，进一步丰富特征图的表达能力，以提高目标的识别和定位准确度，解决误识别问题。与未加入SPP模块的网络进行对比，空间金字塔池化对比实验结果如表2所示。

表2 空间金字塔池化对比实验结果

Table 2 Results of space pyramid pooling ablation experiment

Model	AP/%			Precision/%	Recall/%	mAP/%	FPS
	Bicycle	Car	Person				
YOLOv4-Tiny(original)	32.85	78.65	61.84	**82.69**	41.33	57.78	**232**
YOLOv4-Tiny+SE	38.47	80.79	66.61	80.35	50.18	61.95	**226**
YOLOv4-Tiny+SE+SPP	**42.74**	81.56	68.69	80.54	**53.08**	64.33	212
YOLOv4-Tiny+CBAM	38.97	80.34	66.70	79.65	50.48	62.00	215
YOLOv4-Tiny+CBAM+SPP	42.04	**81.89**	**69.56**	80.33	53.04	**64.50**	202

由表2结果中可以看出，加入SPP模块后，所设计的两种网络对每个类别的AP均有提升。YOLOv4-Tiny+SE+SPP与YOLOv4-Tiny+SE相比，召回率提升2.9%，准确率提升0.19%，平均准确率提高2.38%；YOLOv4-Tiny+CBAM+SPP与YOLOv4-Tiny+CBAM相比，召回率提升2.56%，准确率提升0.68%，平均准确率提高了2.5%。改进的最优网络模型YOLOv4-Tiny+CBAM+SPP相较于原始网络YOLOv4-Tiny，在检测速度上，从232 fps下降至202 fps，减少30 fps；在检测精度方面，mAP从57.78%上升至64.50%，提升6.72%。实验结果表明，通过加入空间金字塔池化模块和注意力机制后，所设计的网络在检

测速度小幅下降的情况下，可以有效提升红外目标检测的平均准确率。

2.2.3 迁移学习对比实验

迁移学习得到的预训练模型有较好的网络初始化权重，可以提高网络的收敛速度。本实验中，在VOC07+12数据集上对网络进行预训练，然后加载主干特征提取网络的权重，在FLIR红外数据集上进行微调训练，与未进行迁移学习的对比实验结果如表3所示。

表3 迁移学习对比实验结果

Table 3 Results of ablation experiments on pre-trained models

Model	AP/%			Precision/%	Recall/%	mAP/%	FPS
	Bicycle	Car	Person				
YOLOv4-Tiny(original)	32.85	78.65	61.84	82.69	41.33	57.78	**232**
① YOLOv4-Tiny+SE+SPP	42.74	81.56	68.69	80.54	53.08	64.33	212
① +Pretraining(VOC07+12)	56.45	84.38	74.38	82.26	61.47	71.74	212
② YOLOv4-Tiny+CBAM+SPP	42.04	81.89	69.56	80.33	53.04	64.50	202
② +Pretraining(VOC07+12)	**61.12**	**84.41**	**75.05**	**83.57**	**63.74**	**73.53**	202

由表3结果中可以看出，通过加入预训练权重策略，可以较大地改善了自行车和人两类的检测准确率。与未加入预训练策略的网络对比，YOLOv4-Tiny+SE+SPP和YOLOv4-Tiny+CBAM+SPP通过加入预训练权重策略进行训练后，准确率分别提高1.72%和3.24%，召回率分别提高8.39%和10.7%，mAP分别提高8.41%和9.03%，各项指标均有明显地提升。实验结果表明，迁移学习是一种较优的训练策略，在FILR红外数据集中，预训练得到良好的初始化权重对模型的收敛速度和模型收敛后的平均准确率均有提升。

2.2.4 嵌入式模型部署实验

本文嵌入式计算平台选择华为的Atlas 200 DK进行网络模型部署实验。华为Atlas 200 DK上搭载了昇腾310 AI计算芯片，其中包含2个DaVinci AI Core和8个A55 Arm Core，半精度(FP16)上最高可以达到11 TFLOPS，典型功耗为20W。

由于在嵌入式端采用了半精度加速推理策略，会对网络权重进行量化，导致网络的准确率有所变化。在本节实验中，针对网络的推理速度和准确率进行对比实验，将原始网络YOLOv4-Tiny和所设计的两种网络YOLOv4-Tiny+SE+SPP与YOLOv4-Tiny+CBAM+SPP部署于华为Atlas 200 DK嵌入式计算平台并进行推理测试，嵌入式模型部署对比实验的结果如表4所示。

表4 嵌入式模型部署对比实验的结果

Table 4 Experimental comparison results of embedded model deployment

Model	Device	AP/%			Precision/%	Recall/%	mAP/%	FPS
		Bicycle	Car	Person				
YOLOv4-Tiny(original)	2 080 Ti	32.85	78.65	61.84	82.69	41.33	57.78	**232**
YOLOv4-Tiny+SE+SPP(our)		56.45	84.38	74.38	82.26	61.47	71.74	212
YOLOv4-Tiny+CBAM+SPP(our)		**61.12**	**84.41**	**75.05**	**83.57**	**63.74**	**73.53**	202
YOLOv4-Tiny(original)	**310 AI**	33.32	78.45	61.05	84.56	40.57	57.60	**82**
YOLOv4-Tiny+SE+SPP(our)		54.61	83.02	72.27	83.15	59.17	69.96	78
YOLOv4-Tiny+CBAM+SPP(our)		**61.87**	**84.15**	**74.61**	**84.88**	**63.46**	**73.54**	71

由表4结果中可以看出，相比于在台式计算机GPU计算资源2 080 Ti，采用嵌入式昇腾310 AI计算芯片时，网络的推理速度大幅降低，网络的平均准确率略有变化。在嵌入式计算平台上，与原始网络YOLOv4-Tiny相比，所提最优网络YOLOv4-Tiny+CBAM+SPP在多方面的性能有着明显提升：在平均准确率方面，从57.60%提升至73.54%，提高了15.94%；在召回率方面，从40.57%提升至63.46%，提高了22.89%；在推理速度方面，从82 fps下降至71 fps，下降了13.41%，但依然可以达到71 fps，满足实时检测。实验结果表明，YOLOv4-Tiny+SE+SPP在嵌入式平台可以达到78 fps的实时检测速度和69.96%的平均准确率，

YOLOv4-Tiny＋CBAM＋SPP在嵌入式平台可以达到71 fps的实时检测速度和73.54%的平均准确率，达到了检测速度和平均准确率的良好平衡，满足嵌入式平台的实时目标检测需求。

2.2.5 网络的性能对比

3类检测目标在不同网络上的P-R曲线，如图6所示，从图中可以看出3类目标的AP均有提升，其中，类别bicycle的AP值有大幅的提升，类别person和car的AP值有一定的提升。与台式计算机GPU计算平台的网络对比，在含有昇腾310AI计算芯片的嵌入式平台上，网络的P-R曲线略微有所变化，平均准确率略微变化。

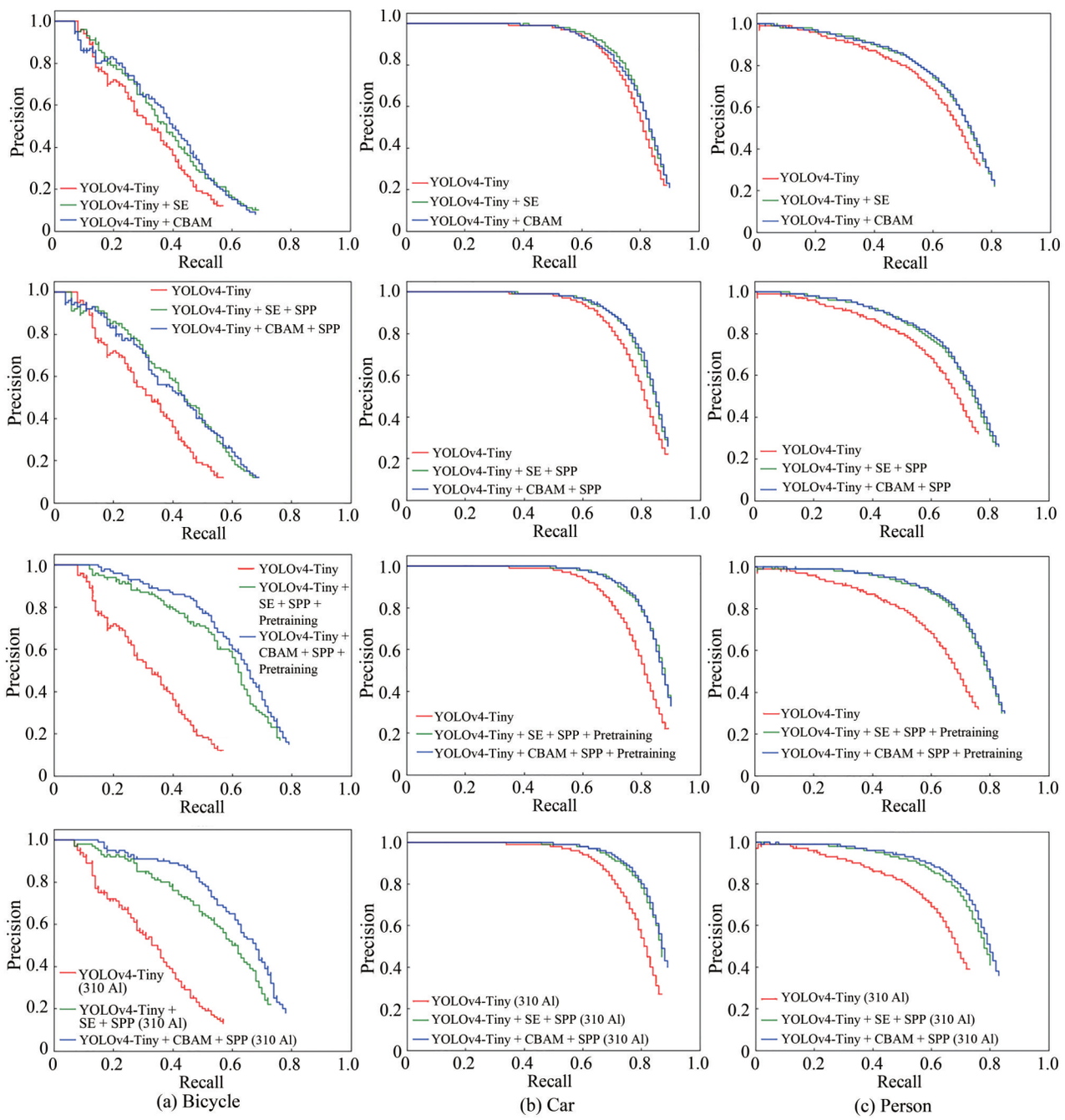

图6 不同类别在不同网络上的P-R曲线

Fig.6 P-R curves of different categories on different networks

2.3 结果分析

四种复杂场景的检测结果如图7所示，采用Grad-CAM在不同网络中的可视化结果如图8所示。从检测结果中可以看出，相较于原网络，所提的两种网络可以有效提高检测能力；从可视化结果中可以看出，通

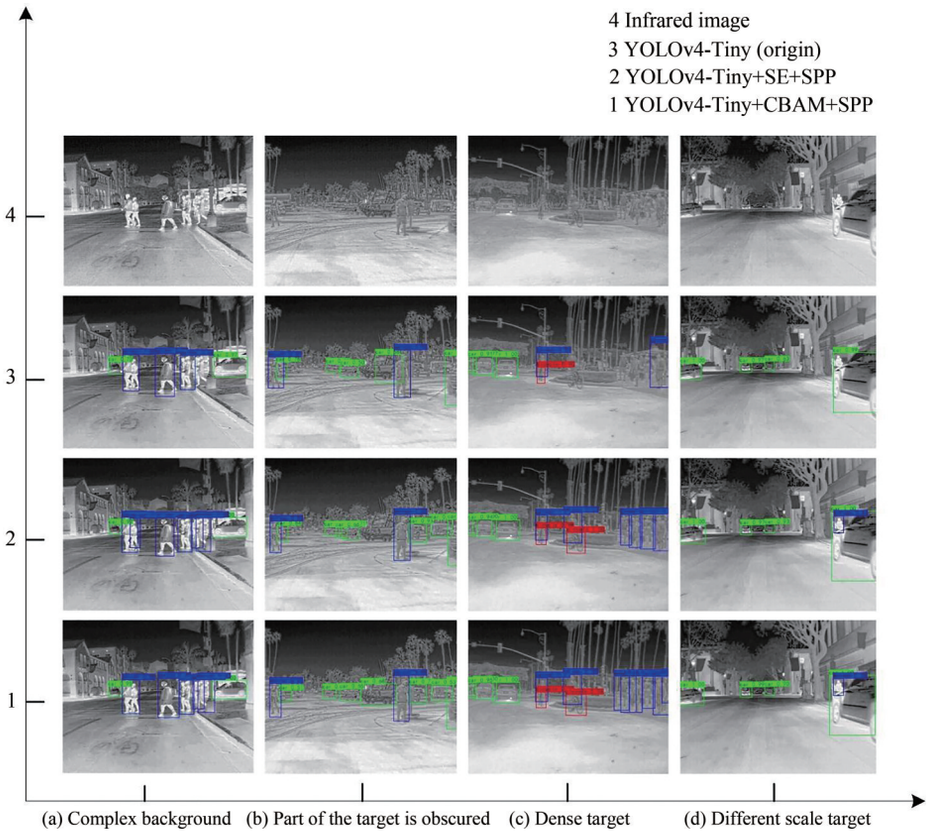

图 7 不同模型检测结果

Fig.7 Detection results of different models

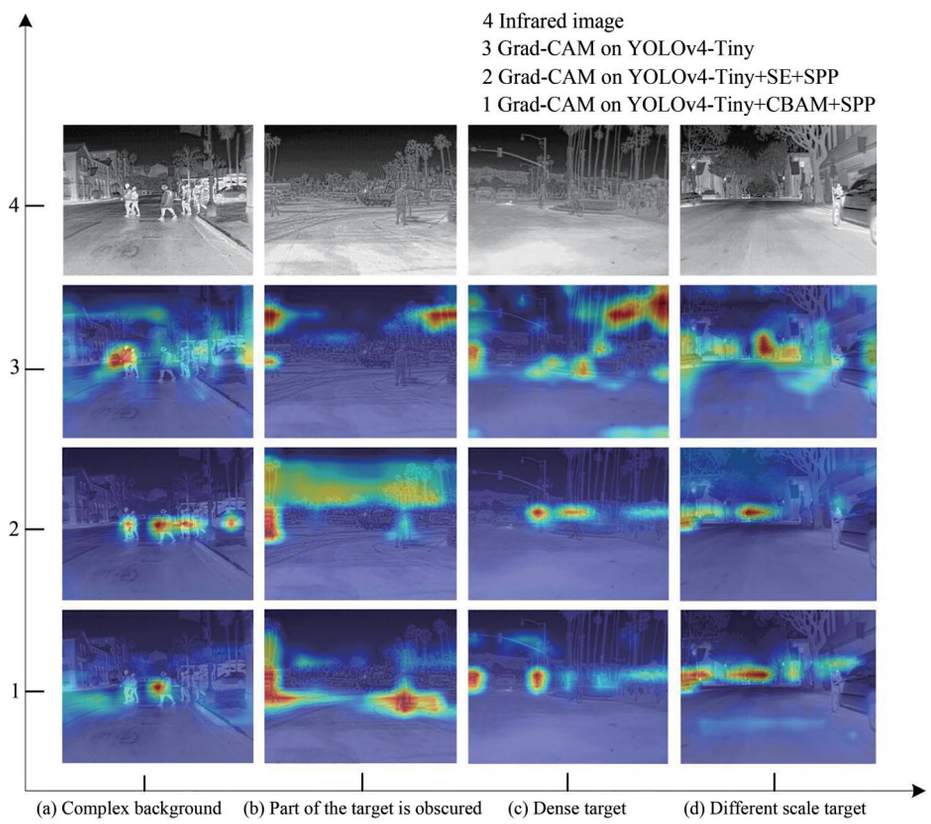

图 8 Grad-CAM 在不同网络中的可视化结果

Fig.8 Grad-CAM in different networks visualization results

过加入注意力机制和空间金字塔池化结构,可以有效增强网络对目标的关注度。场景8(a):与YOLOv4-Tiny相比较,改进后的两种网络可以有效解决复杂背景干扰问题,提高了召回率;场景8(b):YOLOv4-Tiny+CBAM+SPP相较于YOLOv4-Tiny+SE+SPP和YOLOv4-Tiny,在识别数量和定位精度方面中明显占优势,说明了注意力机制CBAM比SE在该场景下,检测效果更佳;场景8(c):改进后的两种网络相较于YOLOv4-Tiny,有效地提高召回率,其中YOLOv4-Tiny+CBAM+SPP表现更佳;场景8(d):改进后的网络可以准确定位尺度不同的目标、识别出遮挡的目标。

3 结论

本文提出两种基于YOLOv4-Tiny的改进网络:YOLOv4-Tiny+SE+SPP和YOLOv4-Tiny+CBAM+SPP,通过融入视觉注意力机制,增强网络对有效特征层的关注度;通过添加SPP模块,对经过注意力加强的特征层进行多尺度特征融合,进一步丰富了特征图的表达能力;利用Grad-CAM在改进的网络中的进行可视化分析,显示改进后的网络可以有效地提高对目标的关注度。实验结果表明,改进的两种网络相较于原网络YOLOv4-Tiny在检测性能方面有较大地提升:在平均准确率方面,YOLOv4-Tiny+SE+SPP提升了13.93%,YOLOv4-Tiny+CBAM+SPP提升了15.75%;在召回率方面,YOLOv4-Tiny+SE+SPP提升了20.14%,YOLOv4-Tiny+CBAM+SPP提升了22.41%;在PC平台上的检测速度,YOLOv4-Tiny+SE+SPP达到了212 fps,YOLOv4-Tiny+CBAM+SPP达到了202 fps。将两种网络模型部署于含有昇腾310AI计算芯片的Atlas 200 DK嵌入式平台上,YOLOv4-Tiny+SE+SPP和YOLOv4-Tiny+CBAM +SPP的推理速度分别可以达到78 fps和71 fps,mAP分别达到69.96%和73.54%,能够兼顾实时性和准确率,可满足红外目标在嵌入式平台的实时检测需求。

参考文献

[1] CHEN Dongjie, ZHANG Wensheng, YANG Yang. Detection and recognition of high-speed railway catenarv locator based on deep learning[J]. Journal of University of Science and Technology of China, 2017, 47(4): 320-327.
陈东杰, 张文生, 杨阳. 基于深度学习的高铁接触网定位器检测与识别[J]. 中国科学技术大学学报,2017, 47(4): 320-327.

[2] ZHANG Zongbao, TAN Zhimin, ZHANG Li, et al. Ship Recognition from infrared images based on deep convolutional neural network[J]. Ship Electronic Engineering, 2020, 40(8): 102-106,165.
刘宗宝, 谭智敏, 张力, 等. 基于深度卷积神经网络的可见光舰船目标识别系统[J]. 舰船电子工程,2020, 40(8): 102-106,165.

[3] XIA Ye, CHEN Limu, WANG Junjie, et al. Single shot multibox detector based vessel detection method and application for active anti-collision monitoring[J]. Journal of Hunan University(Natural Sciences), 2020, 47(3): 97-105.
夏烨, 陈李沐, 王君杰, 等. 基于SSD的桥梁主动防船撞目标检测方法与应用[J]. 湖南大学学报(自然科学版),2020, 47(3): 97-105.

[4] 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), 2016.

[5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]. Springer, Cham, 2016.

[6] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. 2017 IEEE International Conference on Computer Vision (ICCV), 2017: 2999-3007.

[7] WANG Xiaoqing, WANG Xiangjun. Real-time target detection method applied to embedded graphic processing unit[J]. Acta Optica Sinica, 2019, 39(3): 274-280.
王晓青, 王向军. 应用于嵌入式图形处理器的实时目标检测方法[J]. 光学学报,2019, 39(3): 274-280.

[8] JIN Xin, ZHAO Xu, ZHAO Chaoyang, et al. Real-time crowd counting for embedded systems with high accuracy[J]. Chinese High Technology Letters, 2020, 30(1): 32-40.
金鑫, 赵旭, 赵朝阳, 等. 面向嵌入式系统的高精度实时人群计数算法研究[J]. 高技术通讯,2020, 30(1): 32-40.

[9] LU Jian, DENG Bo, Que Longcheng. Stable object tracking method for complex infrared ground environment[J]. Acta Photonica Sinica, 2019, 48(10): 1010001.
吕坚, 邓博, 阙隆成. 复杂红外地面环境下的稳定目标跟踪方法[J]. 光子学报,2019, 48(10): 1010001.

[10] JU Moran, LUO Haibo, LIU Guangqi, et al. Infrared dim and small target detection network based on spatial attention mechanism[J]. Optics and Precision Engineering, 2021, 29(4): 843-853.
鞠默然, 罗海波, 刘广琦, 等. 采用空间注意力机制的红外弱小目标检测网络[J]. 光学精密工程,2021, 29(4): 843-853.

[11] HaimingTIE, FU Guangyuan, LI Shiyi, et al. Typical target detection for infrared homing guidance based on YOLO v3[J]. Laser & Optoelectronics Progress, 2019, 56(16): 147-154.
陈铁明, 付光远, 李诗怡, 等. 基于YOLO v3的红外末制导典型目标检测[J]. 激光与光电子学进展,2019, 56(16): 147-154.

[12] ZHAO Bin, WANG Chunqiang, FU Qiang, et al. Multi-scale infrared pedestrian detection based on deep attention mechanism[J]. Acta Optica Sinica, 2020, 40(5): 41-52.
赵斌, 王春平, 付强, 等. 基于深度注意力机制的多尺度红外行人检测[J]. 光学学报, 2020, 40(5): 41-52.
[13] BOCHKOVSKIY A, WANG C Y, LIAO H. YOLOv4: Optimal speed and accuracy of object detection [C]. Computer Vision and Pattern Recognition, 2020.
[14] JIE H, LI S, GANG S, et al. Squeeze-and-excitation networks [C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017.
[15] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional Block Attention Module [C]. Springer, Cham, 2018.
[16] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(9): 1904-1916.
[17] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization [J]. International Journal of Computer Vision, 2020, 128(2): 336-359.
[18] 适用于算法训练的免费FLIR热数据集[EB/OL]. https://www.flir.cn/oem/adas/adas-dataset-form/.

Real-time Infrared Target Detection Algorithm for Embedded System in Complex Scene

ZHANG Penghui[1], LIU Zhi[2], ZHENG Jianyong[3], HE Boxia[1], PEI Yuhao[1]
(1 *School of Mechanical Engineering, Nanjing University of Science and Technology, Nanjing* 210094, *China*)
(2 *Nanjing Planetech Intelligent Technology Co., LTD, Nanjing* 210014, *China*)
(3 *Institute of Artificial Intelligence, Shanghai University, Shanghai* 200444, *China*)

Abstract: In order to solve the problems of low accuracy and recall rate of infrared target detection under complex background conditions, as well as slow inference speed of network model on embedded computing platform, lightweight network YOLOv4-Tiny was taken as the basic architecture of the algorithm, combined with visual attention mechanism and spatial pyramid pooling idea. Two real-time infrared target detection networks for embedded systems are proposed. Among them, there are a lot of background interference information in target detection in infrared complex scenes. Therefore, the visual attention mechanism is used to effectively learn the weight distribution of the feature map, recalibrate the feature map, strengthen the focus on the target, reduce the influence of irrelevant background information and improve the detection and recognition ability of the model. Spatial pyramid pooling can fuse multi-scale features, enrich the information of feature maps and improve the ability of infrared target recognition and location at different scales. Grad-CAM was used to visualize the feature map strengthened by the attention mechanism, showing the attention of the network model to the target region. The training is carried out on a 2080Ti GPU computer platform using the transfer learning strategy, and deployed on the Atlas 200 DK embedded computing platform with Ascend 310 AI chip as the core. The experimental results show that compared with the original network YOLOv4-Tiny, the infrared images with a resolution of 640 pixels × 512 pixels are detected on the computer platform. The average accuracy and recall rate of the proposed YOLOv4-Tiny+SE+SPP network were improved by 13.96% and 20.14%, respectively, and the inference speed reached 212 FPS. The average accuracy and recall rate of the proposed YOLOv4-Tiny+CBAM+SPP network were improved by 15.75% and 22.41%, respectively, and the inference speed reached 202 FPS. On Atlas 200 DK embedded computing platform, infrared images with a resolution of 640 pixel×512 pixel are detected, compared with the original network YOLOv4-Tiny. The average accuracy and recall rate of the proposed YOLOv4-Tiny+SE+SPP network were improved by 12.36% and 18.6%, respectively, and the inference speed reached 78 FPS. The average accuracy and recall rate of the proposed network YOLOv4-Tiny+CBAM+SPP are improved by 15.94% and 22.89%, respectively, and the inference speed reaches 71 FPS, which can meet the needs of real-time detection and tracking of infrared targets in military and security fields.

Key words: Infared image; Visual attention; Transfer learning; Target detection; Embedded platform
OCIS Codes: 100.4996; 110.3080; 330.7326; 120.1880

Foundation item: National Natural Science Foundation of China (No. 51575281), Fundamental Research Funds for the Central Universities (No. 30916011304)