

基于无预训练卷积神经网络的红外车辆目标检测

陈 皋¹, 王卫华¹, 林丹丹²

(1. 国防科技大学 电子科学学院 ATR 重点实验室, 湖南 长沙 410073; 2. 昆明物理研究所, 云南 昆明 650233)

摘要: 为解决基于卷积神经网络的目标检测算法对预训练权重的过度依赖, 特别是数据稀缺条件下的红外场景目标检测, 提出了融入注意力模块来缓解不进行预训练所带来的检测性能下降的方法。本文基于 YOLO v3 算法, 在网络结构中融入模仿人类注意力机制的 SE 和 CBAM 模块, 对提取的特征进行通道层面和空间层面的重标定。根据特征的重要程度, 自适应地赋予不同权重, 最终提升检测精度。在构建的红外车辆目标数据集上, 注意力模块能够显著提升无预训练卷积神经网络的检测精度, 融入了 CBAM 模块的网络检测精度为 86.3 mAP。实验结果证明了注意力模块能够提升网络的特征提取能力, 使网络摆脱对预训练权重的过度依赖。

关键词: 目标检测; 红外目标; 深度学习; 卷积神经网络

中图分类号: TP391 文献标识码: 文章编号: 1001-8891(2021)04-0342-07

Infrared Vehicle Target Detection Based on Convolutional Neural Network without Pre-training

CHEN Gao¹, WANG Weihua¹, LIN Dandan²

(1. National Key Laboratory of Science and Technology on ATR, School of Electronic Science and Technology,

National University of Defense Technology, Changsha 410073, China;

2. Kunming Institute of Physics, Kunming 650223, China)

Abstract: To tackle the over-dependence of convolutional neural network-based target detection algorithms on pre-training weights, especially for target detection of infrared scenarios under data-sparse conditions, the incorporation of attention modules is proposed to alleviate the degradation of detection performance owing to the absence of pre-training. This paper is based on the YOLO v3 algorithm, which incorporates SE and CBAM modules in a network that mimics human attentional mechanisms to recalibrate the extracted features at the channel and spatial levels. Different weights are adaptively assigned to the features according to their importance, which ultimately improves the detection accuracy. On the constructed infrared vehicle target dataset, the attention module significantly improved the detection accuracy of the non-pre-trained convolutional neural network. Furthermore, the detection accuracy of the network incorporating the CBAM module was 86.3 mAP, demonstrating that the attention module can improve the feature extraction ability of the network and free the network from over-reliance on the pretrained weights.

Key words: target detection, infrared target, deep learning, convolutional neural network

0 引言

通过卷积层、池化层、激活函数等模块层层堆叠的结构设计, 卷积神经网络具有强大的特征提取和拟合能力, 为计算机视觉领域带来了翻天覆地的变化, 已经在图像分类、目标检测、实例分割等任务上取得了优异成绩。目标检测需要在图像中确定目标的位置

和类别, 对应着定位和分类两个子任务。传统方法将多种算法应用于检测任务的不同阶段, 如阈值分割^[1]用于目标定位、梯度直方图特征^[2]用于目标特性描述、支持向量机^[3]用于特征分类。不同于传统的处理思路, 深度学习采用了一体化设计的卷积神经网络, 克服了手工设计特征难度大且适用范围窄的缺陷, 实现了图像输入和检测结果输出的端到端式处理流程, 相

收稿日期: 2020-07-17; 修订日期: 2020-07-30.

作者简介: 陈皋 (1996-), 男, 安徽人, 硕士研究生, 研究方向为红外图像处理、目标检测。E-mail: jzchenriver@163.com.

较于传统方法,检测效果有着显著提升,已经成为当前的主流研究方向。根据网络设计思路的不同,基于卷积神经网络的目标检测算法可分为双阶段和单阶段两类,包括 Faster RCNN^[4]、Thunder Net^[5]、YOLO^[6]、Center Net^[7]等。

双阶段算法也被称为基于分类的检测, Faster RCNN^[4]首先通过区域建议网络在图像上获取目标所在区域的建议,然后根据这些建议区域的特征进行分类和位置优化。Thunder Net^[5]作为面向移动端的实时检测网络,采用轻量化的网络设计,是第一个基于 ARM 平台的实时检测算法。文献[8]利用数据挖掘的方法从生成的 anchor 中筛选出难样本(Hard negative),设计了用于检测航拍图像中车辆的卷积神经网络。单阶段算法也被称为基于回归的检测。以 YOLO^[6]算法为例,整体网络由骨架网络和 YOLO 网络两部分构成,前者用于提取图像的特征,后者根据特征实现对目标位置、类别和置信度的预测,单阶段算法计算量少,检测速度快。Center Net^[7]对于目标的检测不再基于 anchor,而是基于中心点,有效提高了目标遮挡时的检测效果。

深度学习方法需要大量的图像和标注用于训练,训练数据的不足容易导致过拟合,通常会使用 Image Net^[9]、MSCOCO^[10]等大规模数据集预训练的权重进行网络的初始化,再在此基础上用实验数据进行训练。在进行红外目标检测时,样本获取的成本高、难度大,少量的实验样本难以支撑网络的从头训练,需要借助预训练权重来初始化网络。但在红外领域并没有公开的大规模数据集,因此不得不选择可见光图像预训练的权重。然而网络结构与权重是一一对应的,一旦结构发生改变,原本的权重就不再适用,而重新训练来获取预训练权重的代价往往十分昂贵,需要消耗大量的人力物力,也就是说,预训练权重严重制约了卷积神经网络在红外目标检测中的实际应用,为了摆脱这种依赖,本文对无预训练的卷积神经网络展开研究。以红外热像仪拍摄的城市街道场景的车辆为检测目标,基于 YOLO v3^[11]网络进行实验。在设计网络时,为提升网络的特征提取能力,在网络结构中融入注意力模块,包括 SE(squeeze excitation)^[12]和 CBAM(convolution block attention module)^[13],使其能从有限的训练样本中获取更多的有用信息。实验结果表明,改进后网络的检测能力得到大幅提升,结合训练次数的调整,最终实现了对有预训练网络的超越,检测精度可达 86.3 mAP。

1 YOLO v3 算法原理

YOLO 算法将目标检测看作是回归问题,输入图

像后,同时得到目标位置和类别的预测,计算量少,处理速度快。改进版本 YOLO v3^[11]中,加入了 anchor 机制,并采用残差设计加深骨架网络,根据目标的大小在 3 个尺度上进行预测,有效提升了对于小目标的检测精度。

网络深度的加深在一定程度上使网络特征提取的能力得到增强,但在梯度反向传播时,过深的网络容易导致梯度弥散,使训练难以进行。YOLO v3 在设计骨架网络时借鉴了 Res Net^[14]的设计思想,在卷积模块的输入和输出之间加入短接(Shortcut)结构,卷积层输出的是期望特征与输入之间的残差,降低了数据拟合的难度,再利用短接结构与输入相加,最终得到期望的特征输出。该骨架网络包含 53 个卷积层,被称为 DarkNet-53。

为了提升对目标位置预测的精度,YOLO v3 中加入了 anchor 机制。事先对训练集中的目标位置信息进行 K-Means^[15]聚类,即紧密贴合目标的矩形框中心点坐标、宽度和长度,聚类中心作为 anchor 的默认设置。在 YOLO v3 中,一共得到 9 个聚类结果,并根据面积大小分配到 3 个不同的预测分支。面积较小的 anchor 负责对小目标的检测,对应于较为浅层的特征输出;面积较大的 anchor 负责对大目标的检测,对应了深层的特征输出。通过在多个尺度的特征图进行目标检测,大幅提升了小目标检测的精度,补足了原有方法的缺陷。

YOLO 将图像划分为 $s \times s$ 个网格,对应骨架网络输出的大小为 $s \times s$ 的特征图,每个网格负责预测 3 个边界框。网络并不直接预测边界框的位置信息,而是预测边界框与 anchor 之间的偏移量,具体的对应关系如图 1 所示。其中, (t_x, t_y, t_w, t_h) 为网络输出, (c_x, c_y) 为所在网格左上点的坐标, $\sigma(\cdot)$ 为 sigmoid 函数, (p_w, p_h) 为对应 anchor 的宽度和高度, (b_x, b_y, b_w, b_h) 为预测的目标位置坐标。

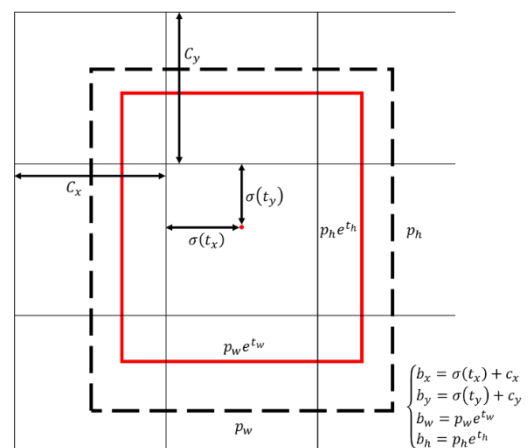


图 1 YOLO v3 中目标位置的预测

Fig.1 The prediction of the target location in the YOLO v3

2 注意力机制

2.1 卷积神经网络中的注意力

在人眼接触某一场景信息时，为了及时准确地做出应对措施，在快速搜索过后，我们的注意力会重点关注场景中的某些区域。在注意力的指导下，后续“资源”会被更高效合理地调度。换句话说，人注意力的分布是不均匀的，并且这种分布会朝着获取更有用信息的方向移动，而这些有用信息会最终服务于人的自然反应或主观意志^[12]。

多数基于卷积神经网络的目标检测算法可以看作是一种编码-解码模型，输入一幅图像进行编码，再从中解码出目标的位置和类别信息作为输出。在这一过程中，可以加入注意力机制。通过对编码信息加权处理，模拟人注意力的不均匀分布，偏向性分配可利用的计算资源，从而获得更加有用的编码信息。训练中，权值分布朝着损失函数下降的方向不断移动，网络学习到不同信息的重要程度，直至收敛，最终完成目标检测任务。

目标检测网络中卷积层输出的特征图即为编码信息，除了宽度和高度外，还有深度，一般的网络对于特征图的所有维度均匀对待。SE^[12]模块在通道维度上添加了注意力机制，对于不同通道的特征图进行权值重标定，学习不同通道的重要程度，突出重要的特征，抑制无用的特征。CBAM^[13]则在通道注意力的基础上，添加了空间注意力，对同一特征图的不同位置进行加权。在注意力模块的帮助下，网络能够提取图像中更具信息量的成分，并最终提升目标检测的精度。

2.2 SE 模块

SE^[12]模块全称 squeeze-and-excitation，包含了压缩 (squeeze) 和激励 (excitation) 两步操作，能够有效增强深度网络的性能。传统卷积层输出的特征图是由各通道直接相加得到的，有学者认为这时通道之间的相关性是隐含的，且卷积核的局部感受野导致了全局信息的忽视。通过 SE 模块中压缩和激励两步操

作，可以提取全局信息，并完成卷积响应的重校准，增强网络对有效特征的敏感度。具体的流程如图 2 所示。

图中， F_{sq} 表示压缩， F_{ex} 表示激活。对于卷积层输出的特征图 U ，($C \times W \times H$) 分别表示深度、宽度、高度，在空间维度上，利用全局平均池化对每一幅特征图进行压缩，得到维度为 ($C \times 1 \times 1$) 的通道描述符 Z ，这样可以提供感受野以外的全局信息，该过程可用下式表示：

$$Z_{avg} = F_{sq}(U) = \text{avg}(U)$$

在激活过程中，包含了 bottleneck 结构的两个全连接层和两个激活函数，可由下式表示：

$$S = F_{ex}(Z_{avg}) = \sigma[f_{c_2}(\delta(f_{c_1}(Z_{avg})))]$$
 (1)

式中： $f_{c_1}(\cdot)$ 表示维度 $C \rightarrow \frac{C}{r}$ 的全连接层； $\delta(\cdot)$ 表示

ReLU 激活函数； $f_{c_2}(\cdot)$ 表示维度 $\frac{C}{r} \rightarrow C$ 的全连接层；

$\sigma(\cdot)$ 表示 sigmoid 激活函数，得到的结果 S 可以认为是不均匀分布的注意力，体现了不同通道的重要程度。通过这些操作，通道之间的相关性被更充分地挖掘出来。最后 S 与 U 相乘，完成通道校准。

2.3 CBAM 模块

CBAM^[13] 模块全称 convolution block attention module，与 SE^[12] 模块相比，对于卷积层输出的特征图，不仅生成一维的通道注意力，同时生成二维的空间注意力，进一步提升网络对有效特征的提取能力。

通道注意力生成的具体步骤如图 3 所示。对于卷积层输出的特征图，在空间维度上，对每一幅特征图进行全局最大池化和平均池化，结果分别输入权值共享的多层感知器，共有 3 层，进行 $C \rightarrow \frac{C}{R} \rightarrow C$ 的维度变化，输出结果的各元素相加，经过激活函数后得到最终的通道注意力。整体流程可由下式表示：

$$Z_{avg} = \text{avg}(U)$$
 (2)

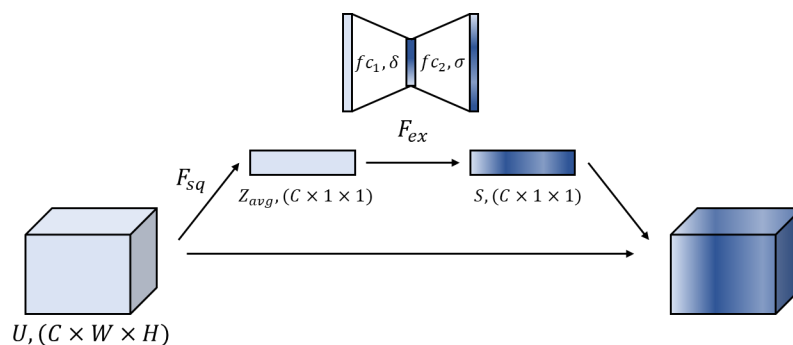


图 2 SE 模块的结构

Fig.2 The structure of SE module

$$Z_{\max} = \max(U) \quad (3)$$

$$S_c = \sigma[f_{c_2}(\delta f_{c_1}(Z_{\text{avg}}))] + f_{c_2}(\delta f_{c_1}(Z_{\max})) \quad (4)$$

式中: $\text{avg}(\cdot)$ 表示全局平均池化; $\max(\cdot)$ 表示全局最大池化; $f_{c_1}(\cdot)$ 表示维度 $\frac{C}{r} \rightarrow C$ 的全连接层; $\delta(\cdot)$ 表示 ReLU 函数; $f_{c_2}(\cdot)$ 表示维度 $\frac{C}{r} \rightarrow C$ 的全连接层; $\sigma(\cdot)$ 表示 sigmoid 函数。

空间注意力生成的具体步骤如图 4 所示。对于卷积层的输出,在通道维度上进行最大池化和平均池化,得到二维的中间结果,将两者在通道维度上相连,输入卷积层,结果经过损失函数,得到单通道二维的空间注意力。整体流程由下式表示:

$$Z_{\text{avg}} = \text{avg}(U) \quad (5)$$

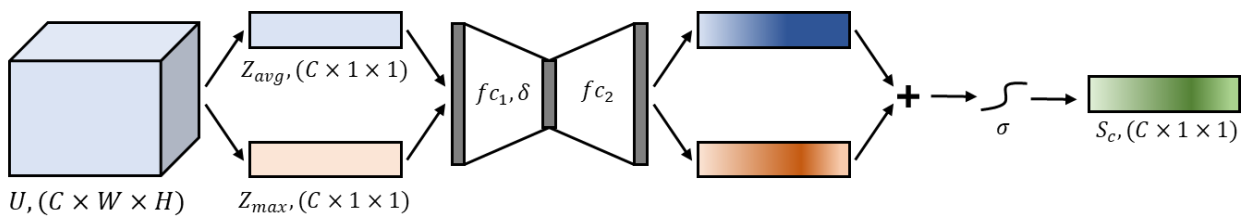


图 3 CBAM 模块中的通道注意力
Fig.3 The channel attention in the CBAM

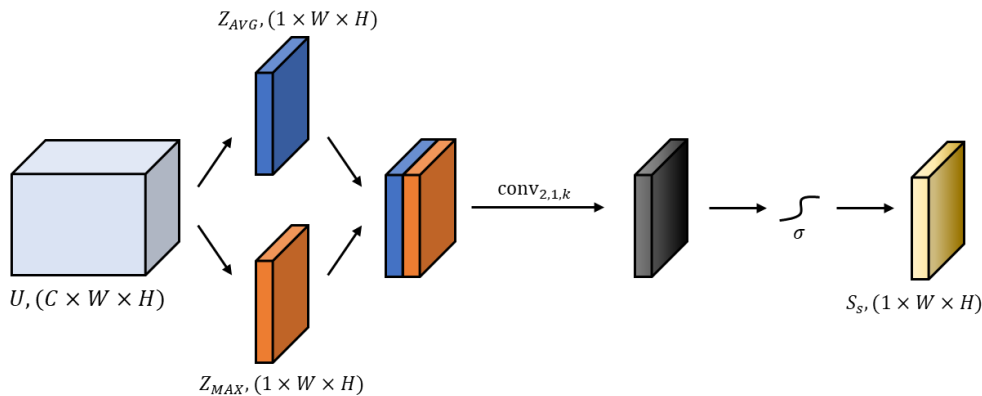


图 4 CBAM 模块中的空间注意力
Fig.4 The spatial attention in the CBAM

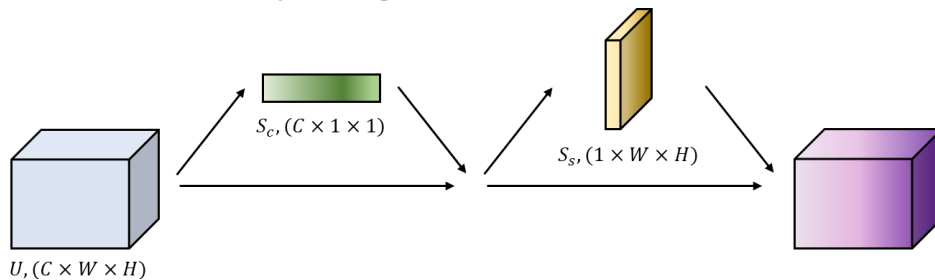


图 5 CBAM 模块的结构
Fig.5 The structure of CBAM

$$Z_{\max} = \max(U) \quad (6)$$

$$S_s = \sigma(\text{conv}_{2,1,k}(Z_{\text{AVG}}; Z_{\max})) \quad (7)$$

式中: $\text{AVG}(\cdot)$ 表示通道维度上的平均池化; $\text{MAX}(\cdot)$ 表示通道维度上的最大池化; $\text{conv}_{2,1,k}(\cdot)$ 表示输入通道为 2, 输出通道为 1, 卷积核为 k 的卷积; $\sigma(\cdot)$ 表示 sigmoid 函数。

通道注意力 S_c 和空间注意力 S_s 先后与 U 相乘得到优化后的输出特征图, 整体流程如图 5 所示。

3 改进的 YOLO v3 网络

卷积神经网络中的注意力模块模仿了人接触外界信息时注意力的不均匀分布, 计算量略微提升, 但能够有效增强网络对有效特征的敏感度。同时, 作为一个即插即用模块, 注意力模块能够十分方便地嵌入已有网络。

预训练权重初始化的 YOLO v3 算法对可见光通用目标有着优越的检测效果,但对于红外目标检测来说,目前没有类似的大规模红外数据集可供预训练,多数情况下,不得不利用可见光数据预训练的权重进行网络初始化。但红外图像与可见光图像的成像机理截然不同,两种图像中目标的特性也有着显著差别,往往需要修改原有的检测可见光目标的网络,从而更好地适应红外场景。但是网络结构与网络权重是一一对应的,一旦根据红外目标检测的特点进行了修改,已有的预训练权重不再适用。如果利用 MSCOCO 数据集重新进行预训练,有着巨大的人力、物力和时间成本。如果直接在实验数据上训练,检测效果将十分依赖数据量的大小,而收集并标注大量的红外数据也是费时费力的。为解决卷积神经网络检测红外目标时遇到的上述困难,我们改进了 YOLO v3 算法,将注意力模块嵌入原始网络中。在不进行预训练的情况下,有效提升网络的特征提取能力,从有限的训练样本中挖掘出更多的有用信息,最终提高检测性能。

如图 6 所示,在骨架网络 DarkNet-53 的每一个残差块中加入注意力模块,3 个预测分支中,最后一个卷积层之前也加入注意力模块。改进后的 YOLO v3 网络结构如表 1、2 所示。

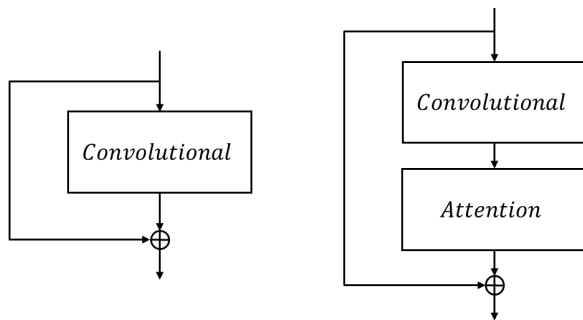


图 6 改进前后的残差块

Fig.6 The residual block before and after improvement

4 实验

4.1 实验设置

实验数据由一台长波红外热像仪拍摄,目标为城市街道场景中的车辆,包括轿车、卡车和公交车 3 类目标。考虑到不同的外界环境,分别在中午 11 点、下午 4 点和晚上 7 点 3 个时间段采集数据,每个时间段的数据包括了大小两个视场,涵盖了目标不同的角度和尺寸。图像分辨率为 1024×768 ,共 630 张,为考察训练样本稀少时的网络性能,不进行数据增广,并按照 7:3 的比例随机划分训练集 441 张,测试集 189

张。

实验平台为 Linux 操作系统,CPU 为 Intel Xeon E5-2678 v3,GPU 为 Nvidia RTX 2080ti,内存 128 G。实验基于 Pytorch,训练的初始学习率为 0.01,随着训练轮次的增加,按照余弦函数进行学习率衰减。采用 SGD (stochastic gradient descent) 算法进行参数更新,动量为 0.937,权重衰减为 0.0005。

表 1 改进的 DarkNet-53 结构

Name	Type	Num	Size
	Convolutional	32	3×3
	Convolutional	64	$3 \times 3, 2$
Stage 1	$\left(\begin{array}{l} \text{Convolutional} \\ \text{Convolutional} \\ \text{Attention} \\ \text{Residual} \end{array} \right) \times 1$	32	1×1
		64	3×3
		128	$3 \times 3, 2$
		64	1×1
Stage 2	$\left(\begin{array}{l} \text{Convolutional} \\ \text{Convolutional} \\ \text{Attention} \\ \text{Residual} \end{array} \right) \times 2$	128	3×3
		256	$3 \times 3, 2$
		128	1×1
		256	3×3
Stage 3	$\left(\begin{array}{l} \text{Convolutional} \\ \text{Convolutional} \\ \text{Attention} \\ \text{Residual} \end{array} \right) \times 8$	512	$3 \times 3, 2$
		256	1×1
		512	3×3
		256	1×1
Stage 4	$\left(\begin{array}{l} \text{Convolutional} \\ \text{Convolutional} \\ \text{Attention} \\ \text{Residual} \end{array} \right) \times 8$	1024	$3 \times 3, 2$
		512	1×1
		1024	3×3
		512	1×1
Stage 5	$\left(\begin{array}{l} \text{Convolutional} \\ \text{Convolutional} \\ \text{Attention} \\ \text{Residual} \end{array} \right) \times 4$	1024	$3 \times 3, 2$
		512	1×1
		1024	3×3
		512	1×1

实验指标包括 IoU 阈值为 0.5 时的 mAP,置信度阈值为 0.4 时的准确率 (Precision) 和召回率 (Recall)。其中 mAP@0.5 为主要指标,综合考虑了所有的置信度阈值。

表 2 改进的预测分支结构

Table 2 The structure of improved prediction subnet

Name	Type	Num	Size
Prediction	$\begin{pmatrix} \text{Convolutional} \\ \text{Convolutional} \end{pmatrix} \times 1$	N	1×1
		$2N$	3×3
$N=512,256,128$			
Attention			
	Convolutional	$(4+1+N_{cls})$	1×1
YOLO			

4.2 实验结果

训练 120 个轮次后,不同网络的实验结果如表 3 所示,使用 MSCOCO 预训练权重进行初始化的 YOLO v3 算法实现了 84.7 mAP。一旦不进行预训练,检测结果会出现剧烈的下降,仅为 38.9 mAP,加入注意力模块后,下降的幅度得到了缓解,说明注意力模块能够提高网络对图像中有效信息的敏感度,帮助提升检测性能,证明了所提方法的合理性。对比两个不同的注意力模块, CBAM 的效果要好于 SE,这归功于空间注意力的加入,进一步提升了网络的检测性能。图 7 展示了不同训练轮次时的测试结果。

表 3 训练 120 轮次的实验结果对比

Table 3 The comparison of results when epoch=120

Epoch=120	Precision	Recall	mAP@0.5
YOLO v3 -mscoco-pretrained	82.3	77.0	84.7
YOLO v3 -no-pretrained	65.3	22.4	38.9
YOLO v3-SE	68.9	36.2	51.7
YOLO v3-CBAM	78.2	40.5	61.7

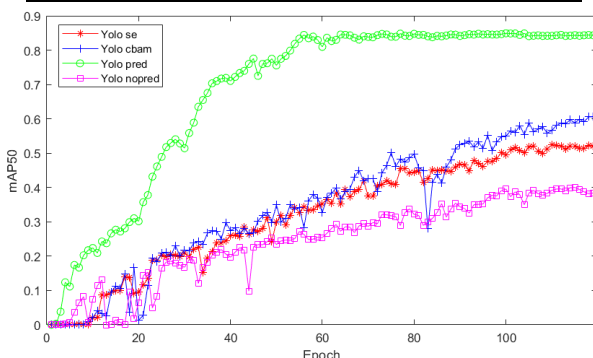


图 7 训练 120 轮次的测试结果

Fig.7 The results on the test dataset when epoch=120

在训练 120 个轮次时,即使加入了注意力模块,不进行预训练带来的检测性能下降也不能被完全消除。因此,将训练轮次调整到 300,实验结果如表 4 所示。

训练 300 个轮次后,在测试集上,3 个不经过预训练的网络的检测表现都有了明显提升,但对于未经改动的 YOLO v3 来说,80.6 mAP 的结果依然要低于预训练的网络。而加入注意力模块后,即使不进行预训练,通过训练次数的增加,网络最终的检测结果为 86.3 mAP,超过了预训练网络。对于两个不同的注意力模块, CBAM 结合了通道注意力和空间注意力,效果要好于只加入了通道注意力的 SE 模块。不同训练轮次时的测试结果如图 8 所示。测试集上的部分检测结果如图 9 所示。

表 4 训练 300 轮次的实验结果对比

Table 4 The comparison of results when epoch=300

Epoch=300	Precision	Recall	mAP@0.5
YOLO v3- no-pre- trained	83.7	65.3	80.6
YOLO v3-SE	87.0	72.4	85.6
YOLO v3 -CBAM	87.8	75.8	86.3

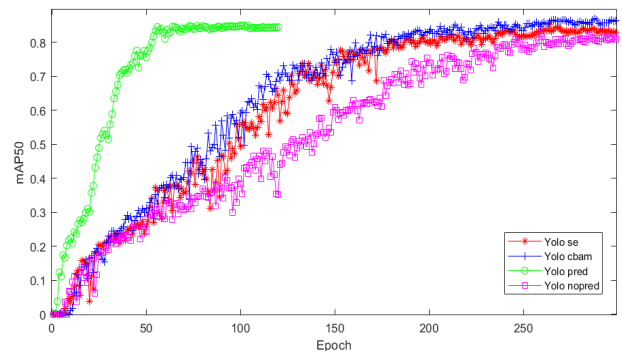


图 8 训练 300 轮次的测试结果

Fig.8 The results on test dataset when epoch=300

5 结论

当卷积神经网络检测红外图像中的目标时,预训练权重扮演着十分重要的角色,但也严重限制了实际的使用。一旦预训练权重对于网络不再适用或不可获得,网络的检测性能会出现大幅的衰减。为摆脱对预训练权重的过度依赖,本文基于 YOLO v3 算法,将 SE 和 CBAM 两种注意力模块嵌入原始的网络中,提高网络对有用特征的敏感度,缓解了不进行预训练所带来的检测性能下降,并通过训练轮次的调整,最终使无预训练网络的检测表现超越了预训练网络,对于城市街道场景中的车辆目标有着良好的检测效果。

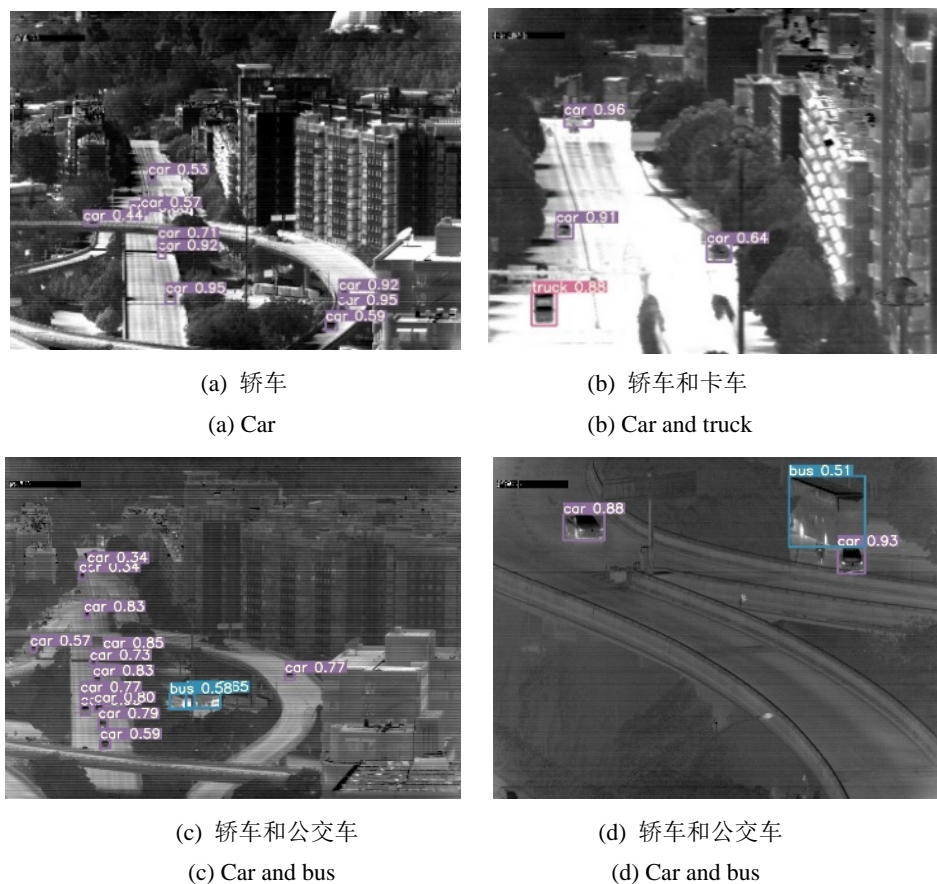


图9 测试集的部分检测结果

Fig.9 Part of detection results on the test dataset

参考文献:

[1] Otsu N. A threshold selection method from gray-level histograms[J]. *IEEE Transactions on Systems, Man, and Cybernetics*, 1979, 9(1): 62-66.

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C]//*Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005: 886-893.

[3] Suykens J A K, Vandewalle J. Least squares support vector machine classifiers[J]. *Neural Processing Letters*, 1999, 9(3): 293-300.

[4] REN S, HE K, 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*, 2015, 39(6): 1137-1149

[5] QIN Z, LI Z, ZHANG Z, et al. ThunderNet: Towards real-time generic object detection on mobile devices[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2019: 6718-6727.

[6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 779-788.

[7] DUAN K, BAI S, XIE L, et al. Centernet: keypoint triplets for object detection[C]//*Proceedings of the IEEE International Conference on Computer Vision*, 2019: 6569-6578.

[8] TANG T, ZHOU S, DENG Z, et al. Vehicle detection in aerial images based on region convolutional neural networks and hard negative example mining[J]. *Sensors*, 2017, 17(2): 336.

[9] DENG J, DONG W, Socher R, et al. Imagenet: a large-scale hierarchical image database[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009: 248-255.

[10] LIN T Y, Maire M, Belongie S, et al. Microsoft coco: common objects in context[C]//*Proceedings of the European Conference on Computer Vision*, 2014: 740-755.

[11] Redmon J, Farhadi A. YOLOv3: An incremental improvement[DB/OL]. <https://arxiv.org/abs/1804.02767>. 2020-0703.

[12] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 7132-7141.

[13] Woo S, Park J, Lee J Y, et al. Cbam: convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany: Springer, 2018: 3-19.

[14] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770-778.

[15] Krishna K, Murty M N. Genetic K-means algorithm[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 1999, 29(3): 433-439.