·开发研究与工程应用·　　文章编号：1000-3428(2022)09-0269-08　　文献标志码：A　　中图分类号：TP391

基于弱语义分割的轻量化交通标志检测网络

曾雷鸣[1,3]，侯　进[2,3]，陈子锐[2,3]，周浩然[1,3]

（1.西南交通大学 计算机与人工智能学院，成都 611756；2.西南交通大学 信息科学与技术学院，成都 611756；
3.西南交通大学 综合交通大数据应用技术国家工程实验室，成都 611756）

摘　要：针对现有网络在检测高分辨率交通标志图片时速度过慢、精确度较低等问题，提出一种轻量化交通标志检测网络。在MobileNetv3-Large基础上对YOLOv4网络的骨干部分进行优化，针对数据集的特点舍弃部分耗时层，更改第8层和第14层的输出通道数，并改进基础模块中通道域注意力网络的注意力机制，使输出的权重数值能更准确地表征特征的重要程度。在检测头前加入基于弱语义分割的动态增强附件，利用其输出作为空间权重分布来矫正激活区域，以避免提取能力下降导致误检、漏检问题，最终构成YOLOv4-SLite网络。采用滑窗剪裁的方法对高分辨率图片进行训练和预测，从而减少训练时间及增加样本的多样性。在TT100K交通标志数据集上的实验结果表明，相较于YOLOv4基准网络，YOLOv4-SLite网络的mAP@0.5仅下降了0.2%，但模型大小减少了96.5%，响应速度提升了227%，精确度与速度的平衡效果达到了预期。

关键词：交通标志检测；YOLOv4网络；轻量化网络；弱语义分割；注意力机制

开放科学（资源服务）标志码（OSID）：

中文引用格式：曾雷鸣，侯进，陈子锐，等.基于弱语义分割的轻量化交通标志检测网络[J].计算机工程，2022，48(9)：269-276，285.

英文引用格式：ZENG L M，HOU J，CHEN Z R，et al. Lightweight traffic sign detection network based on weak semantic segmentation[J].Computer Engineering，2022，48(9)：269-276，285.

Lightweight Traffic Sign Detection Network Based on Weak Semantic Segmentation

ZENG Leiming[1,3]，HOU Jin[2,3]，CHEN Zirui[2,3]，ZHOU Haoran[1,3]

（1.School of Computing and Artificial Intelligence，Southwest Jiaotong University，Chengdu 611756，China；2.School of Information Science and Technology，Southwest Jiaotong University，Chengdu 611756，China；3.National Engineering Laboratory of Comprehensive Transportation Big Data Application Technology，Southwest Jiaotong University，Chengdu 611756，China）

【Abstract】Aiming at the problems of slow speed and low accuracy in detecting high-resolution traffic sign images in existing networks，a lightweight traffic sign-detection network is proposed. On the basis of MobileNetv3-Large，this study optimizes the backbone of a YOLOv4 network，discards some time-consuming layers according to the characteristics of the dataset，changes the number of output channels of layers 8 and 14，and improves the attention mechanism of Squeeze and Excitation Network（SENet）in the basic module，so that the weight value of the output can more accurately represent the importance of the characteristics.This study adds a dynamic enhanced attachment based on weak semantic segmentation in front of the detection header，and uses its output as the spatial weight distribution to correct the active region，to avoid the problem of false detection and missed detection caused by the decline of extraction ability，and finally form a YOLOv4-SLite network.The sliding window clipping method is used to train and predict high-resolution images，to reduce the training time and increase the diversity of samples. The experimental results on the TT100K traffic sign dataset show that，compared with the YOLOv4 benchmark network，the mAP@0.5 of the YOLOv4-SLite network is lost by 0.2%，but the model size is reduced by 96.5%，and the response speed is increased by 227%. The balance of accuracy and speed achieved meets the expectation.

【Key words】traffic sign detection；YOLOv4 network；lightweight network；weak semantic segmentation；attention mechanism

DOI：10. 19678/j. issn. 1000-3428. 0062671

0　概述

随着无人驾驶技术日渐成熟，国内外均已将该技术推进到限定条件下的开放道路测试阶段。交通标志作为真实道路场景中不可回避的识别对象，其识别和定位成为无人驾驶环境感知的研究热点。然而，现实情景中诸如光照条件难以控制、交通标志损毁、褪色等对实时交通标志的检测造成了很大干扰。

基金项目：四川省科技计划项目（2020SYSY0016）。

作者简介：曾雷鸣（1992—），男，硕士研究生，主研方向为深度学习、目标检测；侯　进（通信作者），副教授、博士；陈子锐、周浩然，硕士研究生。

收稿日期：2021-09-13　　**修回日期：**2021-10-26　　**E-mail：**654516667@qq.com

在实际应用中的测试结果表明，基于感兴趣区域（Region of Interest，RoI），使用HOG[1]、Gabor[2]、Haar-like[3]等人工设计特征对RoI进行分类的传统方法不足以应对上述列举的挑战。

得益于硬件尤其是GPU性能的提升，KRIZHEVSKY等[4]搭建了神经网络AlexNet，并在ImageNet挑战赛中夺冠。受此启发，GIRSHICK等[5]将区域卷积神经网络（Regions with Convolutional Neural Networks Features，R-CNN）引入目标检测领域，用一个完整网络的不同分支完成了分类和定位任务。而后，Fast R-CNN[6]、Faster R-CNN[7]完善了以RoI为导向的两次回归算法流程，确立了端到端的、两阶段的检测模型。与此同时，YOLO[8]、SSD[9]等一阶段方法优化了回归过程，以少量的精确度损失获得了检测速度的大幅提升。随后涌现出的语义融合方法FPN[10]、正负样本平衡的方法Focal loss[11]降低了一阶段模型与两阶段模型的精确度差距，使一阶段检测模型逐渐成为研究的主流。

随着深度神经网络的快速发展，诸多关于交通标志检测的算法被研发推出。文献[12]使用统一的神经网络检测高分辨率交通标志，文献[13]使用多重注意力机制相结合的方法提升检测精确度。目前，交通标志检测算法的主要改进方向有探寻更高效的特征提取网络、更有效的语义融合插件、即插即用的tricks等。

本文选取综合性能较优的一阶段网络YOLOv4[14]作为改进和对比的基准，重新设计特征提取网络来替换YOLOv4中原特征提取网络CSPDarkNet53[15]，以降低模型计算量，从而提升响应速度。在此基础上，舍弃MobileNetv3-Large[16]中的部分耗时层，更改第8层和第14层的输出通道数，并改进基础模块中通道域注意力网络（Squeeze and Excitation Network，SENet）[17]的注意力机制，使输出的权重数值能更准确地表征特征的重要程度。此外，设计一种弱语义分割模块，以检测目标的标注信息作为监督，进行前景和背景的分割，得到一个预测的分割掩膜，并将此掩膜作为空间权重分布对特征进行重新标定，以增强本文网络YOLOv4-SLite应对漏检和误检的能力。

1 相关工作

1.1 YOLOv4网络

YOLOv4是YOLO系列[18-19]网络的第4个版本。相较于前两次迭代对样本划分策略和网络结构等瓶颈的突破，YOLOv4并没有进一步提升，而是综合了跨阶段部分连接（Cross Stage Partial-connections，CSP）、路径聚合网络（Path Aggregation Network，PANet）[20]、Mish激活函数[21]、马赛克增强等近年来深度神经网络研究中优秀的子网络、数据增强和训练技巧，形成了“CSPDarkNet+FPAN-SPP+YOLO-Head”的结构。YOLOv4相较于前一个版本极大增强了模型的特征提取、特征融合和特征学习能力，成为现阶段在精确度和速度表现上均十分优异的一阶段目标检测网络。

1.2 MobileNetv3网络

有研究指出，数据在流经神经网络时最耗时的结构是卷积层。文献[22]提出一种适应移动设备的轻量级网络MobileNetv1，设计深度可分离卷积（Depthwise Separable Convolution，DW），通过使用DW卷积替换传统卷积，达到降低模型参数复杂度和运算量的目的，具体的量化公式如式(1)所示：

$$\frac{F_{\mathrm{dw}}}{F_{\mathrm{conv}}}=\frac{1}{C_{\mathrm{out}}}+\frac{1}{K_{\mathrm{size}}} \tag{1}$$

其中：F_{dw}和F_{conv}分别表示可分离卷积与传统卷积的浮点运算数；C_{out}和K_{size}分别表示输出特征的通道数和卷积核的尺寸。

MobileNetv1其后的两次迭代分别融入了倒残差结构和基于通道域注意力（Squeeze and Excitation，SE）机制的SENet，形成MobileNetv3模型。

1.3 语义分割

语义分割任务利用语义标签对全图像素进行逐一分类，从而得到关于语义信息的分割图像。FCN[23]网络的出现使语义分割有了里程碑式的突破，它摒弃了基于滑窗的方法，转而使用一个全卷积的网络，确立了编码-解码的模型架构，克服了区域标注方法的低精确度和低效率缺点。Mask R-CNN[24]以经典的目标检测网络Faster R-CNN为基础，并额外增加一个检测头作为语义分割的输出。通过与分类、回归分支共享深层特征，完成模块化地修改网络，从而将目标检测算法扩展到语义分割任务中。Mask R-CNN在MS COCO[25]公共数据集上的优良性能表明在一个网络中，不同任务可以通过共享深层特征的方式，利用不同的训练标签和损失函数，完成联合学习的任务。

1.4 实时检测与轻量化的趋势

随着目标检测需求的增多，其应用场景和部署平台也越来越多样。嵌入式设备和移动终端要求模型更小巧，自动驾驶、视频检测等应用要求模型拥有更好的实时性。SSD、YOLO系列等成熟且应用广泛的网络日渐不能应对这些状况。近年来，为适应这些新特性，已出现诸多传统算法的改进版本，比如YOLOv4-tiny[26]和MobileNet-SSD[27]采用简化特征提取网络压缩模型，从而提高实时性。

2 YOLOv4网络改进

2.1 网络整体结构

YOLOv4-SLite网络的整体框架：以改进的MobileNetv3-Lite作为基础语义特征提取网络，以FPN加PAN的组合来实现双重特征融合，最后以SSB弱语义分割模块作为检测头的动态增强附件。网络输入尺寸为512×512，经过32倍、16倍和8倍的下采样倍率得到3种不同尺寸的特征，在网络neck部分进行2次上采样实现浅层特征和深层语义特征的融合，3个检测头的输出分别为64×64×141、32×32×141、16×16×141，分别对应小型目标、中型目标和大型目标的检测。输出中第3个维度的构成如下：每个网格预测3个锚框（anchor），每个anchor包含42个类别值、4个坐标相关值以及1个置信度值（x，y，w，h，conf，Ccls）。当使用R_DP模块时，基础结构中包含残差结构。当使用DP模块时，基础结构中没

有残差结构，用虚线加以区分。YOLOv4-SLite网络的结构如图1所示。

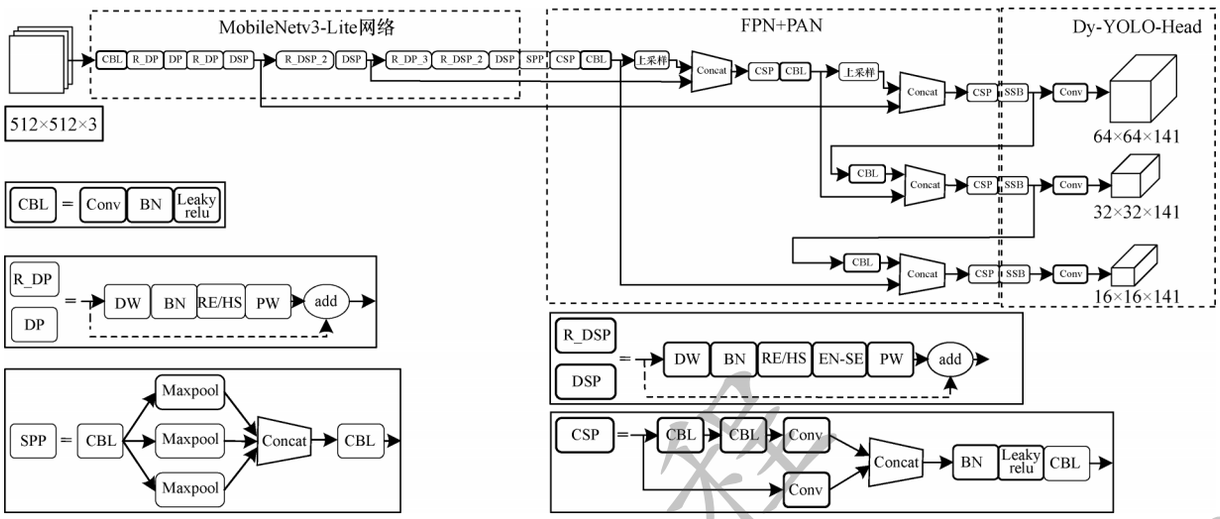

图1　YOLOv4-SLite网络整体结构

Fig.1　Overall structure YOLOv4-SLite network

2.2　改进的特征提取网络MobileNetv3-Lite

MobileNetv3-Large作为一款优秀的轻量化网络，在ImageNet分类、MS COCO检测等基准任务上表现良好。但本文所用交通标志数据集中图片具有高分辨率和目标区域小的独特性，针对这2个特点，本文对MobileNetv3-Large进行了相应改进。

为解决图片裁剪操作拖慢模型检测速度的问题，本文对MobileNetv3-Large进行优化。具体做法：保留原模型的前14层，舍弃其余结构；为避免引入更多卷积来对齐特征融合的输入通道数，将第8层的输出通道数由80更改为40，最后一层的通道数由160更改为112。为抑制特征通道的减少对模型学习能力的剧烈影响，且不过多增加浮点运算数，对MobileNetv3中的重要模块SENet进行改进增强，从而得到MobileNetv3-Lite网络，其结构如表1所示。其中“√”表示使用SE注意力模块，“—”表示不使用SE注意力模块。

表1　MobileNetv3-Lite网络结构

Table 1　Structure of MobileNetv3-Lite network

输入尺寸	运算模块	中间层通道数	输出通道	注意力模块	激活函数	步距
$512^2\times3$	CBL	—	16	—	HS	2
$256^2\times16$	R_DP	16	16	—	RE	1
$256^2\times16$	DP	64	24	—	RE	2
$128^2\times24$	R_DP	72	24	—	RE	1
$128^2\times24$	DSP	72	40	√	RE	2
$64^2\times40$	R_DSP	120	40	√	RE	1
$64^2\times40$	R_DSP	120	40	√	RE	1
$64^2\times40$	DSP	240	40	√	HS	2
$32^2\times40$	R_DP	200	80	—	HS	1
$32^2\times80$	R_DP	184	80	—	HS	1
$32^2\times80$	R_DP	184	80	—	HS	1
$32^2\times80$	R_DSP	480	112	√	HS	1
$32^2\times112$	R_DSP	672	112	√	HS	1
$32^2\times112$	DSP	672	112	√	HS	2

原始的SENet模块通过将输入特征在空间维度上进行压缩，使原本的特征通道转变成一个能表征全局感受野的一维向量，计算原理如式(2)所示：

$$Z_c=\frac{1}{H_1\times W_1}\sum_{i=1}^{H}\sum_{j=1}^{W}f(i,j) \tag{2}$$

其中：$f(i,j)$代表特征图上每个像素点的值；H_1、W_1分别代表特征图的长与宽；Z_c为网络输出，代表该通道重要性的实数。

但SENet模块使用全局平均池化(Global Average Pooling，GAP)作为压缩机制，对于目标占比均衡的任务，通道特征的平均值能够较好地代表该通道的响应情况，因此通过该方法获取全局的上下文关系。正如上文提到，本文的实验数据集具有目标区域占比小的特点，此时输入图片在经过卷积进行特性提取后，在特征图上只能占据很小的激活区域。因此，在空间维度进行特征平均池化会致使背景和干扰信息淹没前景，导致网络对前景目标的响应失真。这是因为目标小，一般作为纹理提取的全局最大池化(Global Max Pooling，GMP)会筛选出信号最强的区域，能够更好地反映该通道对前景目标的响应情况，计算原理如式(3)所示：

$$z_c'=\max_{i,j\in H_1,W_1} f(i,j) \tag{3}$$

基于此，本文重新设计了SENet的压缩机制，使用2个实数表征全局感受野，得到EN-SENet模块。具体实现：分别使用GAP和GMP对每个通道的特征进行压缩，并将池化结果在通道方向上进行拼接，得到维度为$1\times1\times2C$的向量，然后将该向量送入全连接网络，得到通道注意力权重A_c。计算原理如式(4)所示：

$$F=(\sigma(A_c)+1)\times F_1 \tag{4}$$

得到融合通道注意力的特征图F的具体流程如图2所示。

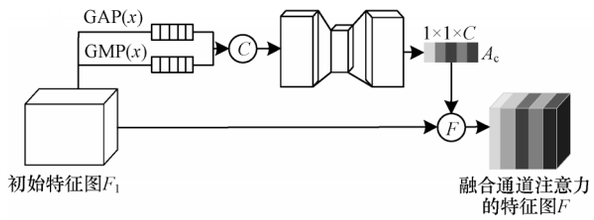

图2 改进的通道注意力模块 EN-SENet

Fig.2 Improved channel attention module EN-SENet

2.3 基于弱语义分割的空间注意力机制

2.3.1 弱语义分割网络

对神经网络的输出进行研究发现，特征图的激活区域与原图中的检测区域存在映射重叠，这说明神经网络在数据流动过程中会逐渐收敛关注范围，将视阈聚焦在真实的检测区域。然而在本文的检测任务中，目标物体占图像的像素面积较小，而且图像多余部分不是纯粹的背景。这些噪声使深层特征表征的检测物体空间映射发生畸变，若将学习到的特征直接输入检测头会严重影响网络对目标的识别与检测效果。文献[28]提出一种解决办法：采取无监督的方式，通过堆叠如图3所示的残差注意力模块，增强特征图中的有效部分。但该方法的缺点也很明显，即整体的运算时间会增加。

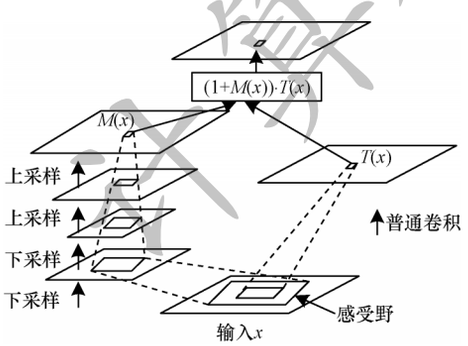

图3 无监督注意力掩码模块

Fig.3 Unsupervised attention mask module

本文旨在设计一种轻量化的检测网路，若沿用这种基础模块层层堆叠的方式，明显与主旨相悖。受文献[13]和文献[28]的启发，本文设计了一种弱语义分割模块，并把它作为一种动态增强附件而非构建网络基础模块，仅在YOLOv4-SLite网络的检测头之前使用一次，将该模块的输出作为模型学习到的空间权重分布来矫正特征，从而提升模型的检测能力。

基于弱语义分割的空间注意力的算法流程如图4所示。将输入图像经MobileNetv3-Lite和特征融合子模块后的输出作为弱语义分割模块的输入特征图，并将输入特征图传入语义分割模块进行前景和背景的预测，从而得到空间注意力权重。接着，把注意力权重分配给对应的输入特征图以加强目标特征。

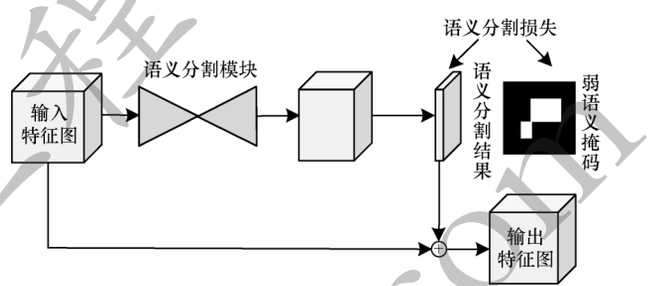

图4 语义分割算法流程

Fig.4 Procedure of semantic segmentation algorithm

2.3.2 语义分割模块

语义分割模块作为一种典型的编码-解码模型，编码部分一般使用卷积或者池化用于缩小特征图尺寸，解码部分采用双线性插值的方法逐级恢复特征图。本文设计了基础语义分割模块SSA和效果增强的语义分割模块SSB，分别如图5和图6所示。

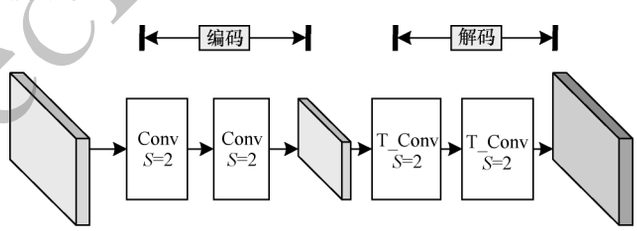

图5 语义分割模块 SSA

Fig.5 Semantic segmentation module SSA

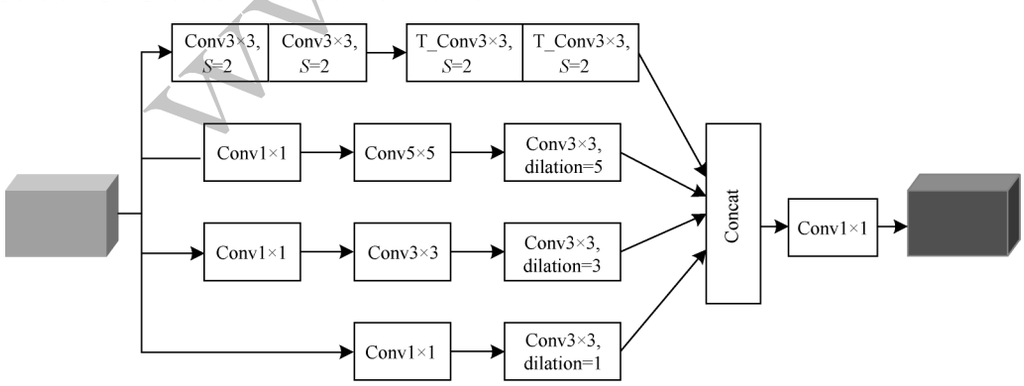

图6 语义分割模块 SSB

Fig.6 Semantic segmentation module SSB

SSA模块参照文献[24]的设计思想，使用语义分割任务中最简单的框架，仅使用2个步距为2的卷积作为编码器，2个步距为2的转置卷积作为解码器，以此来验证语义分割是否有效。为扩大模块感受野的同时结合多尺度的上下文，将经典语义分割模型DeepLabv3中的空洞空间卷积池化金字塔(Atrous Spatial Pyramid Pooling，ASPP)[29]改进后引入SSA模块，最终扩充形成SSB模块。

ASPP模块采用多分支并联的形式，通过膨胀率不同的空洞卷积(Dilated Convolution，DC)获取更大的感受野，再对各分支的特征进行融合，从而获得精确的上下文信息。本文在ASPP模块的基础上进行改进：为降低模块的运算量，对每个分支使用1×1的标准卷积进行降维；将DC层的膨胀率分别降低为1、3、5；对膨胀率为3和5的分支添加3×3的标准卷积来获取基础特征；对4分支的特征进行拼接，最后经过1×1的标准卷积进行特征融合和通道降维。

2.3.3　监督信息和注意力权重的生成

由于本次任务提供的数据集缺少语义标签，因此本文只能利用标注信息来生成语义掩膜，为区别于传统方法，将之称为弱语义掩膜[30]。具体做法：输入原始图片，若语义分割模块输出矩阵上的像素点落入坐标框内，则将该点处的像素数值置1，反之则置0。为避免模糊性，将落在标注框上点的像素数值设置为255，并在训练时忽略掉。然后将该像素数值图映射至与YOLOv4-SLite网络检测头匹配的尺寸，本文没有使用如PASCAL VOC数据集中用过的调色板模式，而是简单选择二值作为区分，得到最终的弱语义掩码。弱语义掩码生成流程如图7所示。使用弱语义掩码进行监督训练，模块SSA或SSB的输出$\boldsymbol{M}_x$即为空间注意力权重。为了防止网络性能的退化，引入残差结构作为恒等映射，即：

$$F_3=(1+\boldsymbol{M}_x)F_2 \tag{5}$$

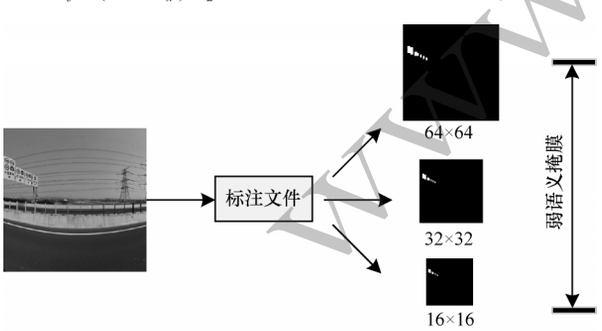

图7　弱语义掩码生成流程

Fig.7　Generation process of weak semantic mask

由于本文只需要语义分割模块进行前景和背景的预测，实际上是个二分类任务，因此只需要计算分割结果$\boldsymbol{M}_x$和弱语义掩膜$\boldsymbol{M}^*$的交叉熵损失，如式(6)所示：

$$L_{\text{mask}}=-\sum_{i}^{h}\sum_{j}^{w}M_{ij}^{*}\log_a(M_{ij})+\alpha(1-M_{ij}^{*})\log_a(1-M_{ij}) \tag{6}$$

其中：h与w分别为弱语义掩膜的长度和宽度；M_{ij}为特征图上坐标为(i,j)的点的像素值；M_{ij}^{*}为掩码上对应坐标的像素值。

3　实验结果与分析

3.1　数据分析与预处理

本文采用清华大学与腾讯公司联合制作的TT100K[12]数据集，数据来源于车载摄像头的街景实拍，具有场景真实、类型多样的特点。该数据集包含训练集图片6 107张，验证集图片3 073张，涵盖了不同光照强度和气候条件下的样本[13]，数据集中的图像分辨率为2 048×2 048像素，其中尺寸小于32×32的检测对象占总目标个数的40.5%，造成了小目标检测的难题。此外，高分辨率图像包含了更多的背景信息，网络更易出现误检情况。本文通过统计数据集中221个类别的分布情况，发现样本分布存在严重的长尾效应，如图8所示。综上所述，相比于其他公开的交通标志数据集，TT100K数据集的挑战难度更大。为抑制长尾效应对模型的影响，本文对数据集进行重构。具体做法：对于数量少于50的类别，直接剔除它们的标注信息；对于数量大于50但小于100的类别，则使用随机高斯噪声、水平翻转、马赛克增强等手段将其样本数扩充至100。最后的数据集包括42个类别，数据分布如图9所示。

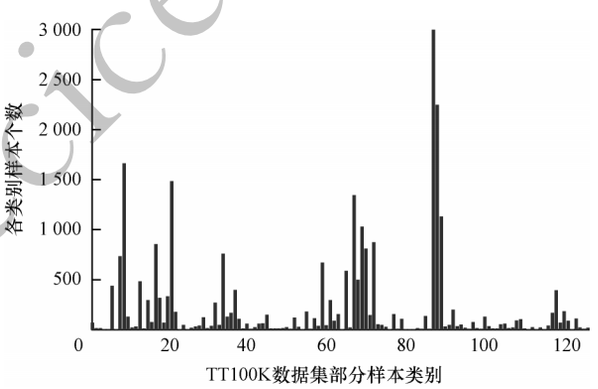

图8　TT100K数据集原始样本分布

Fig.8　Original sample distribution of TT100K data set

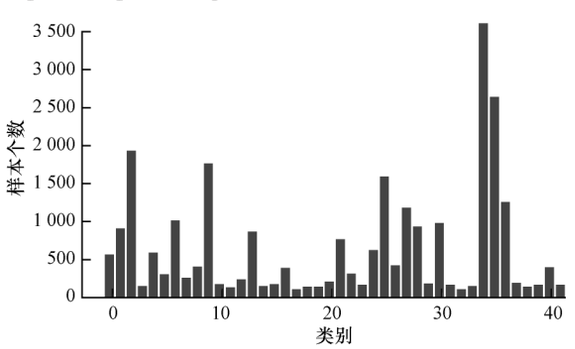

图9　抑制长尾效应后的样本分布

Fig.9　Sample distribution after suppressing the long tail effect

3.2 训练策略

经过3.1节的处理,数据分布不均的问题得到解决,但重构数据集的图像分辨率依旧为2 048×2 048,若直接送入模型进行训练,模型会因显存爆炸而无法训练。若设置过小的批量(batch size)勉强训练,不但会增加训练时间,而且由于相邻batch size差异过大,容易导致模型震荡不能收敛。

为解决这个问题,本文参考文献[31]处理高分辨率图片的方法。使用大小为512×512的滑窗,将重叠率(overlap rate)设置为0.2,对训练集进行裁剪,随机地对图像进行HSV空间颜色变换。滤除不含目标和包含不完整目标的样本之后,共得到15 702张训练图片。通过对高分辨率图片进行裁剪,不但解决了上述提到的问题,而且增加了样本的多样性。

由于增加了语义分割模块进行联合训练,YOLOv4-SLite网络除了回归损失、置信度损失和分类损失之外,还需要额外加上分割损失,总损失的表达式如式(7)所示:

$$L_{\text{total}}=\lambda_1 L_{\text{reg}}+\lambda_2 L_{\text{mask}}+\lambda_3 L_{\text{conf}}+\lambda_4 L_{\text{cls}} \tag{7}$$

其中:λ_1、λ_2、λ_3、λ_4分别为各项损失设置的权重系数。

实验以预处理后的数据集为输入,以深度学习框架Pytorch、Intel Core i7和NVIDA GTX1080Ti搭建运行环境。使用SGD优化器,并使用自定义调度器动态调整学习率,初始化学习率为0.01,动量为0.937,权重衰减为5×10^{-4},batch size设置为48,总共训练300个epoch。

3.3 评价指标

本文使用平均精确度(Mean Average Precision,mAP)、召回率、精确度、浮点运算数共4项指标来评价模型性能,其中精确度与召回率为主要评价指标。但是单一地使用精确度或者召回率指标都不能准确反映模型的性能,为全面评价模型,本文使用主流检测方法,将P-R曲线作为核心指标来验证模型的检测效果。通过计算阈值为0.5时的平均精确度(mAP@0.5),即各个类别AP的平均值,从而定量比较模型的性能。mAP的计算公式如式(8)所示:

$$\text{mAP}=\frac{1}{N}\sum_{0}^{N}\int_0^1 P_n(r)\mathrm{d}r \tag{8}$$

其中:N为该数据集的样本类别数量;P_n为该样本类别的AP值。

3.4 消融实验

3.4.1 改进有效性分析

为验证EN-SENet结构的有效性,将对照组压缩机制设置为SENet,而其他结构与改进组保持一致。检验模型预测是否正确的条件设置如下:预测框与真实框(Ground Truth,GT)的IOU阈值为0.5,类别置信度阈值为0.6。实验结果表明,使用GMP加GAP的拼接组合能显著增强特征提取能力,而浮点运算数仅增加了0.4×10^9 frame/s,具体数据见表2。

表2 不同特征压缩方法的结果对比

Table 2 Comparison of results of different feature compression methods

通道注意力模块	精确度	召回率	浮点运算数/(10^9 frame·s^{-1})
SENet	0.893	0.916	4.6
EN-SENnet	0.915	0.907	5.0

此外,针对本文提出的压缩机制和两种分割模块,本文进行了相关消融实验。分别将MobileNetv3-Lite和语义分割模块SSA和SSB添加到YOLOv4网络中,保持训练和测试方法一致,结果如表3所示,其中"√"表示使用了相应的模块,"—"则表示不使用对应的模块。

表3 模块消融实验

Table 3 Module ablation experiment

EN-SENet	SSA	SSB	精确度	召回率	浮点运算数/(10^9 frame·s^{-1})
√	—	—	0.915	0.905	5.0
√	√	—	0.929	0.911	11.3
√	—	√	0.953	0.927	14.9

通过分析表3数据可知,分割模块能够对检测结果起正面作用,且仅使用基础分割模块SSA就能显著提升模型性能。在综合考量性能提升和计算负担的因素下,增强模块SSB的表现更优。

3.4.2 预测结果分析

通过模块消融实验确定模型使用的模块后,将验证集输入模型,并统计模型预测,结果如表4所示,可以发现:数据集中所有类别(42个)的预测精确度与YOLOv4基本持平。对复杂背景和极小物体的检测情况进行分析,也可以达到预期,预测结果如图10、图11所示。

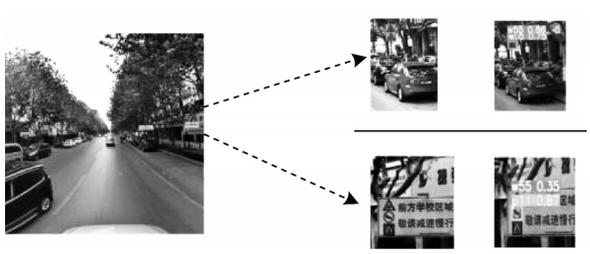

图10 复杂背景下的预测结果

Fig.10 Forecast results in a complex context

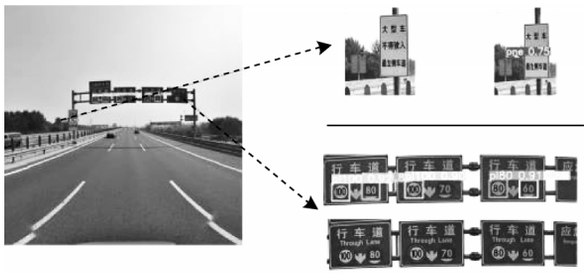

图11　对小目标的预测结果

Fig.11　Predicted results of small targets

表4　YOLOv4-SLite与YOLOv4网络在不同预测类上的预测精确度对比

Table 4　Comparison of prediction accuracy between YOLOv4-SLite and YOLOv4 on different prediction classes

预测类别	YOLOv4网络	YOLOv4-SLite网络
i2	0.974	0.974
I4	0.974	0.973
I5	0.974	0.977
Il100	0.977	0.989
Il60	0.982	0.983
Il80	0.988	0.985
Io	0.940	0.945
Ip	0.985	0.984
P10	0.955	0.958
P11	0.962	0.961
P12	0.981	0.982
P19	0.965	0.969
P23	0.976	0.977
P26	0.959	0.963
P27	0.991	0.992
P3	0.962	0.919
P5	0.979	0.978
P6	0.867	0.854
Pg	0.980	0.984
Ph4	0.927	0.930
Ph4.5	0.981	0.963
Pl100	0.974	0.975
Pl120	0.968	0.970
Pl20	0.978	0.976
Pl30	0.987	0.976
Pl40	0.976	0.978
Pl5	0.967	0.960
Pl50	0.978	0.977
Pl60	0.984	0.977
Pl70	0.960	0.949
Pl80	0.981	0.981
Pm20	0.979	0.963
Pm30	0.982	0.932
Pm55	0.970	0.974
Pn	0.976	0.972
Po	0.917	0.917
Pr40	0.980	0.978
W13	0.964	0.987
W55	0.944	0.974
W57	0.958	0.974
W59	0.891	0.892
Pne	0.978	0.979

3.5　与其他网络的性能对比

本文选取RetinaNet50[11]、YOLOv4[15]、YOLOv3[20]、YOLOv4-tiny[26]、Cascade R-CNN[32]、ATSS[33]、PAA[34]共7种网络进行性能对比，囊括了近几年涌现的经典框架和最新的改进。RetinaNet50通过缓解正负样本不均衡的问题，极大提升了模型性能，成为后续一阶段算法沿用的设计模式；ATSS则沿用RetinaNet框架的例子，证明了影响模型性能的关键因素是正负样本的选择策略，而与回归的方式、anchor的数量无关；PAA进一步消除了ATSS中的超参数，采用概率的方式分配正负样本，进一步提升了模型的性能；Cascade R-CNN通过为多检测头设置不同的IOU改善了性能，是两阶段算法的代表。

将上述模型的P-R曲线绘制在同一坐标轴下，如图12所示，可以清晰看到代表YOLOv4-SLite（本文网络）的曲线包裹了其网络法曲线，并几乎与YOLOv4曲线重合，直观上断定本文网络胜过最新的主流网络，与YOLOv4效果相当。

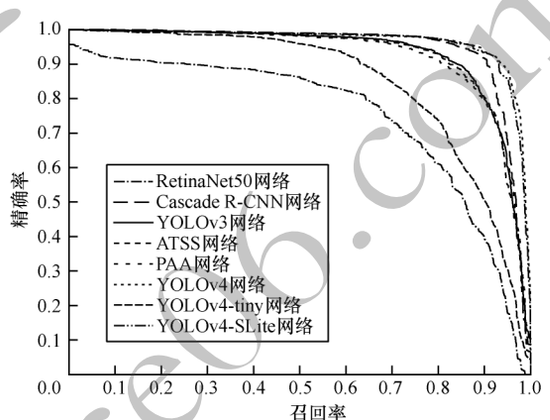

图12　YOLOv4-SLite与其他主流网络的P-R曲线对比

Fig.12　Comparison of P-R curve between YOLOv4-SLite and other mainstream networks

为更全面、综合地比较本文网络的优势，在精确度的基础上，加入模型大小和检测速度两项指标对这7种网络进行对比，结果如表5所示，可以看到，YOLOv4-SLite在精确度和速度综合考虑的情况下明显优于其他网络。此外，相较于YOLOv4网络，YOLOv4-SLite网络在mAP@0.5仅损失0.2%的情况下，模型大小降低了96.5%，浮点运算数提升了227%。

表5　YOLOv4-SLite与其他主流网络的综合比较

Table 5　Comprehensive comparison between YOLOv4-SLite and other mainstream networks

网络	mAP@0.5/%	浮点运算数/(10^9 frame·s^{-1})	模型大小/MB
RetinaNet50网络	85.1	3.2	296.9
Cascade R-CNN网络	94.8	1.7	553.7
YOLOv3网络	93.3	3.4	255.9
ATSS网络	92.8	3.4	257.2
PAA网络	93.1	2.2	253.7
YOLOv4网络	96.6	3.7	243.0
YOLOv4-tiny网络	88.9	16.3	23.8
YOLOv4-SLite网络	96.4	12.1	8.5

4 结束语

针对现有交通标志识别网络在处理高分辨率输入时存在检测速度慢、识别精确度偏低等不足，本文设计一种基于YOLOv4的轻量化改进网络YOLOv4-SLite。为应对轻量化可能导致的性能下降问题，额外为检测头设计一种基于弱语义分割的动态增强附件，且不过多增加模型的浮点运算数。实验结果表明，YOLOv4-SLite相较于YOLOv4基准网络，在mAP@0.5仅损失0.2%的情况下，模型大小降低了96.5%，浮点运算数提升了227%，在综合性能上具有较大优势。但本文仍存在裁剪环节为算法瓶颈以及没有解决YOLO系列网络检测头耦合的问题，而解决这些问题能够提升YOLO系列网络的准确度[35]。下一步将寻求合适的裁剪方式破除瓶颈，并将分类与回归任务进行解耦，以提升模型的性能。

参考文献

[1] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2005: 886-893.

[2] LEE T S. Image representation using 2D Gabor wavelets[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996, 18(10): 959-971.

[3] VIOLA P, JONES M. Rapid object detection using a boosted cascade of simple features[C]//Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2001: 54-62.

[4] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[5] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 580-587.

[6] GIRSHICK R. Fast R-CNN[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 1440-1448.

[7] 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, 2017, 39(6): 1137-1149.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 779-788.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD single shot multibox detector[C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2016: 121-129.

[10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 936-944.

[11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2999-3007.

[12] ZHU Z, LIANG D, ZHANG S H, et al. Traffic-sign detection and classification in the wild[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 2110-2118.

[13] 郭璠，张泳祥，唐琎，等. YOLOv3-A：基于注意力机制的交通标志检测网络[J]. 通信学报，2021，42(1)：87-99.
GUO F, ZHANG Y X, TANG J, et al. YOLOv3-A: a traffic sign detection network based on attention mechanism[J]. Journal on Communications, 2021, 42(1): 87-99. (in Chinese)

[14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. [2021-08-01]. https://arxiv. org/abs/2004. 10934.

[15] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSPNet: a new backbone that can enhance learning capability of CNN [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Washington D. C., USA: IEEE Press, 2020: 1571-1580.

[16] HOWARD A, SANDLER M, CHEN B, et al. Searching for MobileNetv3[C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 1314-1324.

[17] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[C]//Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence. Washington D. C., USA: IEEE Press, 2018: 2011-2023.

[18] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6517-6525.

[19] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. [2021-08-01]. https://arxiv. org/abs/1804. 02767.

[20] WANG W H, XIE E Z, SONG X G, et al. Efficient and accurate arbitrary-shaped text detection with pixel aggregation network [C]//Proceedings of IEEE/CVF International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2019: 8439-8448.

[21] MISRA D. Mish: a self regularized non-monotonic neural activation function[EB/OL]. [2021-08-01]. https://arxiv. org/abs/1908. 08681.

[22] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications[EB/OL]. [2021-08-01]. https://arxiv. org/abs/1704. 04861.

[23] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]// Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2021: 3431-3440.

（下转第285页）

（上接第276页）

[24] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN [C]//Proceedings of IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 2980-2988.

[25] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: common objects in context [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 740-755.

[26] JIANG Z C, ZHAO L Q, LI S Y, et al. Real-time object detection method based on improved YOLOv4-tiny [EB/OL]. [2021-08-01]. https://arxiv.org/abs/2011.04244.

[27] LI Y T, HUANG H S, XIE Q S, et al. Research on a surface defect detection algorithm based on MobileNet-SSD [J]. Applied Sciences, 2018, 8(9): 1678-1683.

[28] WANG F, JIANG M Q, QIAN C, et al. Residual attention network for image classification [C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2017: 6450-6458.

[29] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2018: 833-851.

[30] 周勇, 陈思霖, 赵佳琦, 等. 基于弱语义注意力的遥感图像可解释目标检测[J]. 电子学报, 2021, 49(4): 679-689.
ZHOU Y, CHEN S L, ZHAO J Q, et al. Weakly semantic based attention network for interpretable object detection in remote sensing imagery [J]. Acta Electronica Sinica, 2021, 49(4): 679-689. (in Chinese)

[31] VAN ETTEN A. You only look twice: rapid multi-scale object detection in satellite imagery [EB/OL]. [2021-08-01]. https://arxiv.org/abs/1805.09512.

[32] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 6154-6162.

[33] ZHANG S F, CHI C, YAO Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2020: 9756-9765.

[34] KIM K, LEE H S. Probabilistic anchor assignment with IoU prediction for object detection [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2020: 355-371.

[35] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021 [EB/OL]. [2021-08-01]. https://arxiv.org/abs/2107.08430.

编辑　赖玉玲