文章编号　1004-924X(2023)06-0950-12

采用长距离依赖和多尺度表达的轻量化车辆检测

荆修平，田　莹*

(辽宁科技大学 计算机与软件工程学院，辽宁 鞍山 114000)

摘要：基于深度学习的车辆检测在众多领域发挥着至关重要的作用，是近年来计算机视觉的一个重要发展方向。车辆轻量化检测包含了对网络结构和计算效率的探索，并在智慧交通等诸多领域都得以广泛应用。然而在诸多场景下存在相机中车辆目标尺度变化大、车辆相互遮挡等问题，这些情况会影响到网络检测车辆的精度。针对上述问题，提出改进Yolov5s的车辆检测方法。首先通过视觉注意力网络捕获长距离依赖，对原有特征图施加新的权重，增强自适应性，提升网络的抗遮挡能力；接着在残差模块内部再次构造水平方向残差，在一个模块内部构建相同数量、不同大小感受野的特征图，丰富网络的多尺度表达能力。实验结果表明：改进后的网络在Pascal VOC车辆数据集上提供2.1%mAP性能提升，在MS COCO车辆数据集上提供1.7%mAP性能提升。改进后网络的多尺度表达能力更加出色，且抗遮挡能力更强，与原始网络相比检测结果更具有竞争力。

关　键　词：计算机视觉；车辆检测；轻量化网络；长距离依赖；多尺度表达

中图分类号：TP391　　**文献标识码**：A　　**doi**：10.37188/OPE.20233106.0950

Lightweight vehicle detection using long-distance dependence and multi-scale representation

JING Xiuping，TIAN Ying*

(*College of Computer Science and Software Engineering*，*University of Science and Technology Liaoning*，*Anshan* 114000，*China*)

* *Corresponding author*，*E-mail*：*astianying*@126.*com*

Abstract：Vehicle detection based on deep learning plays a vital role in many fields. In recent years，it has presented a major development direction for computer vision. Lightweight vehicle detection includes the exploration of network structure and computing efficiency，and it is widely used in many fields such as intelligent transportation. However，challenges exist in different scenarios，such as large changes in vehicle scale in detection cameras and vehicles overlapping each other，which reduce the precision of the network in detecting vehicles. To solve these problems，this study proposes an improved YOLOv5s method for detecting vehicles. First，the study proposes to capture long-distance dependencies between objects through a visual attention network and apply new weights to the network's original feature map to increase the adaptability of the network. These operations improve the anti-occlusion ability of the network. Second，the horizontal residual is constructedagain in the residual module. The output feature maps contain the same number and different sizes of receptive fields per module. Feature extraction occurs at a more fine-

收稿日期：2022-06-09；**修订日期**：2022-07-26.

基金项目：国家自然科学基金资助项目(No. 62072086)；辽宁省教育厅资助项目(No. LJKZ0310)

grained level, thereby enriching the multi-scale representation ability of the network. The experimental results show that the improved network provides 2.1% mAP performance on the Pascal visual object classes (VOC) vehicle telemetry dataset and a 1.7% mAP performance on the MS COCO vehicle telemetry dataset. The performance of the improved network is more powerful and its anti-occlusion ability is enhanced. Compared with the original network, the detection results are more competitive.

Key words: computer vision; vehicle detection; lightweight network; long-distance dependence; multi-scale representation

1　引　言

随着人工智能的发展和人们对安全问题的关注，车辆检测[1-2]成为计算机视觉领域的研究热点。车辆检测作为智慧交通的重要组成部分，在自动驾驶、车辆监控等领域得到广泛应用。

传统车辆检测依靠HOG（Histogram of Oriented Gradient）[3]，SVM（Support Vector Machines）[4]等算法，传统算法的鲁棒性较差、泛化能力不足，且人工提取特征的方式过程繁琐复杂，依赖工程人员的自身经验，算法的应用局限性较大，难以应对如今复杂的交通场景。

如今基于深度学习的目标检测算法凭借着强大的特征提取能力和鲁棒性受到了许多研究者的关注。单阶段网络[5-6]不再依赖区域搜索网络生成的候选框，而是在网络最后直接预测坐标信息，只进行一次回归，通常单阶段网络相比于双阶段拥有更强实时性，车辆检测方法随之取得了重大进展。李经宇[7]等人提出一种基于Yolov3车辆检测算法，使用空间金字塔池化模块对多尺度局部区域特征进行融合和拼接。秦鸿睿[8]等人提出一种基于改进EfficientDet的车辆检测方法，首先通过采用马赛克图像增强丰富数据集的多样性，然后采用AdamW优化器替换SGD提高模型训练中收敛效率，最后采用EFL根据正负样本比例及稀有样本的类别动态调节损失贡献。

但单阶段依然存在模型体积大、运算量高且难以在计算资源有限的平台上部署等问题。对于这些平台，计算成本和存储空间至关重要，尤其是在自动驾驶、车辆监控等实时性要求较高的任务中。车辆轻量化检测旨在减轻计算机存储以及运算量负担，包含了针对网络结构和计算效率的探索，推动了计算机视觉等技术在众多设备的应用落地，并在智慧交通等诸多领域都得以广泛应用。轻量化网络不仅可以较为准确感知和理解周围态势，还可以及时掌握路况状态并与外界环境信息实现交互，实现二者权衡，应用场景更加成熟。国内学者做出许多研究工作，张洋[9]提出一种基于改进YOLOv3-tiny的轻量级车辆检测方法，对主干网络进行了分析与改进，模型中引入了CBAM（Convolutional Block Attention Module）注意力机制，分别在通道尺度和空间尺度进行注意加权，并通过K-means获取实验数据集的锚框。张云强[10]等人提出一种基于改进轻量化网络Yolov4-tiny车辆检测方法，对特征金字塔网络第二尺度输出层的最后一个卷积单元输出特征与网络中第二个卷积单元输出特征进行融合，增强网络精度。赵璐璐[11]等人提出一种基于轻量化网络Yolov5s的改进车辆目标检测模型，将注意力机制SE模块引入YOLOv5s网络中，增强网络对车辆重要特征并抑制一般特征以强化检测网络对车辆目标的辨识能力，并引入Focal Loss加强对难分正负样本的学习。

然而在智慧交通诸多场景下存在相机快速运动、尺度变化大、车辆相互遮挡等问题，这些情况会影响到网络检测车辆的精度。基于此，本文提出基于轻量化网络Yolov5s的车辆检测方法，通过视觉注意力网络捕获长距离依赖提升网络的抗遮挡能力，在残差结构内部再次构造水平方向残差以丰富网络的多尺度表达能力。该轻量化网络能够更加精准地检测出车辆位置和车辆信息。

2　相关算法介绍

Yolo（You Only Look Once）[12-14]系列算法是单阶段类型网络，算法发展至今网络日渐成熟。Yolov5作为Yolo系列算法之一，网络模型和算

法思想得到广泛关注和应用。

Yolov5s作为Yolov5轻量化程度最高的网络，拥有较小的模型体积以及优越的推理速度。在训练前预处理阶段，图片首先会经过马赛克图像增强，将图片的长宽随机剪裁，然后拼接在一起，接着会将图片统一缩放到640×640，用灰色填补空缺位置。马赛克增强增加了数据集的多样性，变相增加批标准化大小，有利于加快网络收敛速度。Yolov5s网络整体分为特征提取、特征融合以及预测头三个部分。特征提取采用与CSPDarkNet[15]相似的网络结构设计，分为4个阶段逐阶段提取特征。特征融合采用路径聚合网络Pan[16]结构，将高维的特征图和低维的特征图按通道维度进行连接，用以融合不同尺度的特征信息。特征提取和特征融合部分的残差结构均采用CSP[17]结构，将特征图分割为两个部分，对其中一条分支的特征进行复用，另一条分支重复进行残差特征提取，从而节省网络开销。空间池化金字塔(Spatial Pyramid Pooling, SPP)[18]模块只会在特征主干提取网络最后出现，空间池化金字塔模块首先经过一个降维卷积单元，然后分别经过5×5，9×9，13×13三个不同大小的池化核，然后将三个经过池化后的特征图和经过降维卷积单元的特征图按通道维度进行连接，最后再次经过一个降维卷积单元进行降维，该模块主要负责在高维时获取不同大小感受野的特征图。最后三个预测特征层分别预测三种尺度不同大小的特征图。网络的深度和宽度设计采用和EfficientNet[19-20]系列相同的设计理念，通过制定超参数以此来控制通道倍率和网络深度，实现不同大小的网络模型。

3　改进措施

3.1　长距离依赖

自注意力[21]最初的设计初衷是用来解决自然语言处理方向的问题，但近一年来自注意力出现在视觉领域的各个方向中。在ViT(Vision Transformer)[22]提出后，视觉领域中愈加注重长距离依赖的重要性。

图1为长距离依赖示意图(彩图见期刊电子版)，蓝色深浅表示该特征点的重要性。网络可根据当前特征点较大范围内的特征信息，利用其他特征点自适应加权表示当前特征点本身，反应每个特征点的相关程度。不再受限于较小感受野，在学习中网络会判定未被遮挡区域特征点的二元关系，在面对遮挡情况，网络拥有更强大的抗遮挡能力。

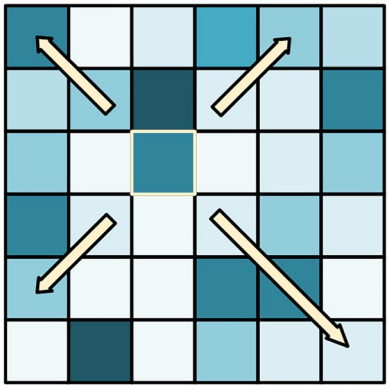


图1　长距离依赖示意图

Fig. 1　Sketch map of long distance dependence

在处理高维特征时采用视觉注意力网络(Visual Attention Network, VAN)[23]捕获长距离依赖。视觉注意力网络由大核注意力(Large Kernel Attention, LKA)和卷积前馈网络(Convolutional Feedforward Network, CFFN)两个子网络构成。图2为视觉注意力网络结构示意图。

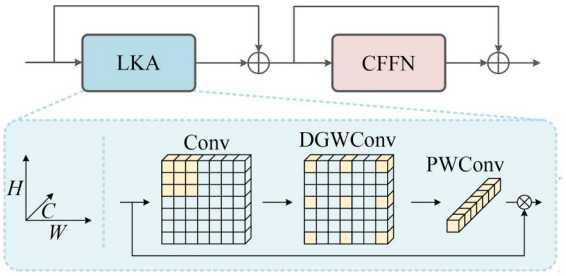


图2　视觉注意力网络结构示意图

Fig. 2　Structure of Visual Attention Network

由图2所示(彩图见期刊电子版)，大核注意力由一个普通卷积(Convolution，Conv)、一个空洞组卷积(Dilated GroupWise Convolution，DGWConv)和一个逐点卷积构成(PointWise Convolution，PWConv)，黄色方块表示卷积过程，忽略补零填充。大核注意力通过超大卷积核的空洞组卷积和逐点卷积来捕获长距离依赖，并且大核

注意力具有线性复杂度。大核注意力计算过程可表示为：

$$\boldsymbol{X}'=PWConv\left(DGWConv\left(Conv\left(\boldsymbol{X}\right)\right)\right)\otimes\boldsymbol{X},\quad(1)$$

其中：$\boldsymbol{X}$为输入特征图，且$\boldsymbol{X}\in\boldsymbol{R}^{c\times h\times w}$，$\boldsymbol{X}'$为大核注意力输出的特征图。$\otimes$为矩阵广播乘法，对输入特征图施加新权重，增加了网络的自适应性。

大核注意力复杂度分析：在计算过程中为了方便简化，对卷积层中的偏置项进行了忽略。其中，输入特征图和输出特征图的通道数相等且为C，K为卷积核大小，d为空洞率，为权衡网络计算量和参数量分组数为4，那么大核注意力的参数量可计算为：

$$\text{Params}=\left\lceil\frac{K}{d}\right\rceil^2\times\frac{C^2}{4}+(2d-1)^2\times\frac{C^2}{4}+C^2,\quad(2)$$

其中：C为常数项，且空洞率d和卷积核大小为奇整数，当$K=21$时，参数量公式(2)可取极小值，因此当空洞率$d=3$时，超大卷积核空洞组卷积的卷积核大小为7。当$K\neq21$时，公式(2)取极值时，空洞率d非奇且无法被除尽，当K过小时长距离依赖捕获能力遭到大幅削弱，当K过大时网络参数量过于庞大。

将大核注意力得到的特征图经过卷积前馈网络，以此来保持网络的非线性能力。卷积前馈网络由两个普通逐点卷积，一个逐深度卷积和一个GELU激活函数组成。特征图依次经过一个逐点卷积、逐深度卷积、激活函数和驻点卷积。两个子网络都会与前一层的输出残差相加。

3.2　多尺度表达

许多由多尺度单元构建的主干网络在各种视觉任务中实现了最先进的性能，例如Zhang H[24]等人提出根据SE(Squeeze and Excitation)[25]注意力权重选择不同大小卷积核的分支。Li D[26]等人提出在一个卷积内部周期循环使用不同空洞率的卷积核。Duta I C[27]等人提出在残差内部构造以垂直方向构造不同尺度的特征信息。虽然这些先进的工作为视觉领域开拓新方向，但本文采用更加合适的方式实现多尺度表达。

目前大多数的方法都是依赖不同大小的卷积核构造多尺度特征，或是通过逐步加深网络得到更多更大感受野的特征图，然而Res2[28]模块则是在一个残差模块内部，以水平残差的方式分层实现多尺度表达。Res2模块在残差内部再次残差，能够以更细粒度的水平表达特征，并且在残差内部增强感受野。图3为Res2模块将3×3卷积单元拆分成4组特征子集结构示意图。

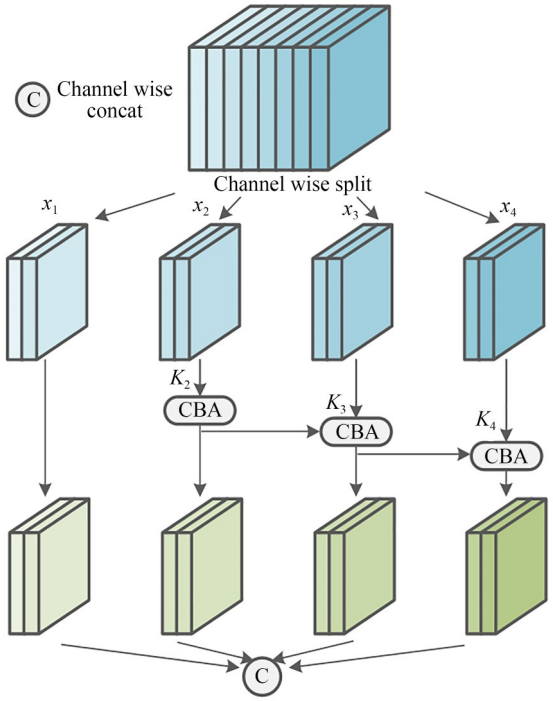


图3　Res2模块4组特征子集结构示意图

Fig. 3　Structure of the Res2 module when feature maps subset are 4

在经过第一个1×1降维卷积单元后，特征图将被等分割为s个特征图子集，特征图子集由$\boldsymbol{x}_i$表示，其中$i\in\{1,2,\cdots,s\}$。每个特征图子集都具有相同的通道数。除了$\boldsymbol{x}_1$所在分支以外，每个$\boldsymbol{x}_i$所在分支都有一个对应的3×3卷积单元，这个卷积单元记为K_i。用$\boldsymbol{y}_i$记为K_i的输出。对于$\boldsymbol{x}_1$所在分支，对这个分支进行特征重用，没有K_i卷积单元，且不与$\boldsymbol{x}_2$所在分支相加。每个特征图子集$\boldsymbol{x}_i$与K_{i-1}的输出相加(其中$i\geqslant3$)，作为下一个水平分支的输入。最终将各个$\boldsymbol{y}_i$连接，得到Res2模块最终输出。因此，$\boldsymbol{y}_i$可记为：

$$\boldsymbol{y}_i=\begin{cases}\boldsymbol{x}_i, & i=1\\ K_i\left(\boldsymbol{x}_i\right), & i=2\\ K_i\left(\boldsymbol{x}_i+\boldsymbol{y}_{i-1}\right), & 2<i\leqslant s\end{cases}.\quad(3)$$

每个K_i都会从对应的特征子集中提取特征，且作为下个水平分支的输入，所以K_i输出特征图的感受野会随i的增大依次递增。Res2模块的输

出包含相同数量和不同大小感受野的特征图。

3.3 网络细节

图 4 为原始 C3 模块结构示意图，特征图输入后会被左侧和右侧的 b_1 和 b_2 分支进行第一次降维，通道的缩减倍率为 0.5，当特征图进入 Bottleneck 残差模块中会被 1×1 的降维卷积单元进行第二次降维。n 为残差模块重复进行次数。两条分支的特征图最后汇总，以通道维度连接在一起。

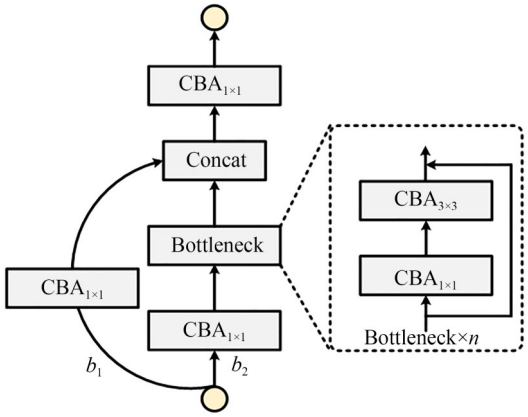


图 4　原始 C3 模块结构示意图

Fig. 4　Structure of original C3 module

原始 Yolov5s 残差模块由一个 1×1 降维卷积单元和一个 3×3 卷积单元构成。在设计 Res2 残差模块时，模块输入和输出特征图的通道数必须相等，否则水平分支无法相加。新的残差模块由一个 1×1 降维卷积单元，一个 Res2 模块，和一个 1×1 升维卷积单元构成。对于每个 Res2 残差模块特征图子集通道数，有不同划分方式，划分方式如表 1 所示。

表 1 中的 Setting 代表不同的 Res2 残差模块特征图子集通道划分方式，#Params 和 FLOPs 分别为不同划分方式 Yolov5s 网络的计算量和参数量。其中 org 为原始网络，w 为控制通道的系数，s 特征图子集分割数。$13w\times4s$ 表示特征图子集分割数为 4，且通道系数为 13 的 Res2 残差模块特征图子集通道分割方式。每个特征图子集的通道数可计算为：

$$c_{x_i}=\left\lfloor c_{-}\cdot\left(w/c_{\text{stage1}}\right)\right\rfloor, \tag{4}$$

其中：c_{-} 为 C3 残差模块中 1×1 降维卷积单元输出通道数，c_{stage1} 为第一阶段 C3 模块第一次降维后的通道数。权衡网络推理速度，实验中选择 $13w\times4s$ 为特征图子集通道数，对原有 C3 的残差模块进行改造。除此以外，改进后的 Res2 残差结构中激活函数由原始的 ReLU 一并被统一为非线性能力更强的 SiLU，且改进后的 Res2 残差结构中的卷积和批标准化可以在推理预处理时进行融合，用以加快网络推理速度。

表 1　不同的特征图子集划分方式

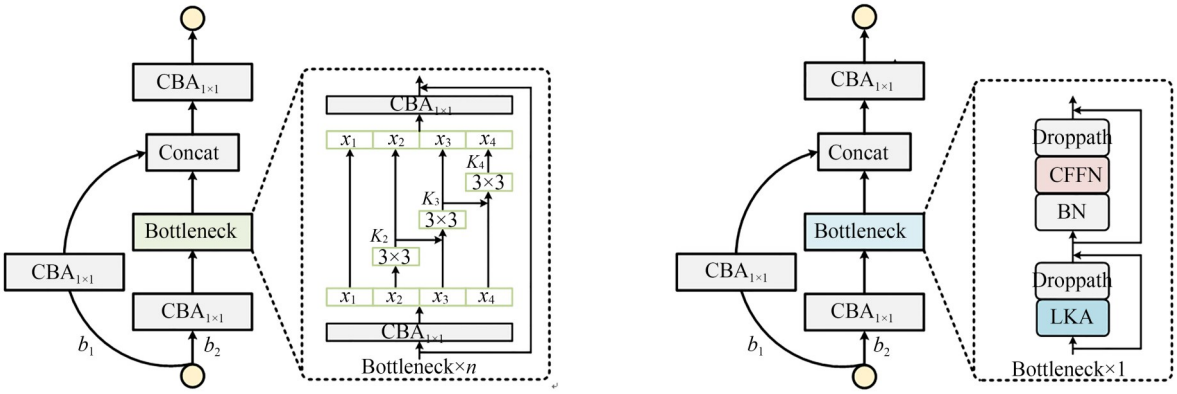
Tab. 1　Different feature map subsets division methods

Setting	#Params(M)	FLOPs(G)
org	7.3	17.1
$7w\times8s$	7.2	16.2
$13w\times4s$	7.3	16.6
$13w\times6s$	7.6	18.6

对于大核注意力中的三个卷积，均为普通卷积，无批标准化和激活函数。除此以外对于视觉注意力网络中的卷积前馈网络，在实验中只提供单倍非线性映射。实验中用 Res2 残差模块和视觉注意力网络对原始 C3 的残差结构进行改造，改进后的 C3 结构分别命名为 Res2C3 和 VANC3。图 5(a)和(b)分别为改进后的 Res2C3 和 VANC3 的结构示意图。

VANC3 只出现在特征提取最后一个阶段，且和原网络一样只迭代一次。Res2C3 将构建在阶段 1 至阶段 3。Res2C3 在低纬度构建多尺度特征，VANC3 在高维度捕获长距离依赖。

图 6 为 Yolov5s 改进后的网络结构示意图。主干网络特征提取分为 4 个阶段，且每个阶段特征图的下采样倍率逐阶段增加，在第 4 阶段最高下采样倍率为 32 倍。Focus 模块负责将特征图每四个特征点分为一组，每组中位置相同的特征点互相拼接在一起，最后将拼接后的特征图按通道维度连接，DP(Down Sampling)为步长为 2 的普通卷积单元，这两个模块均负责提供 2 倍下采样。普通卷积单元包含一个普通卷积、批标准化和 SiLU 激活函数。Res2C3 和 VANC3 为改进后的 C3 模块。为权衡网络综合性能，并未对路径聚合网络做出任何改动。



(a) Res2C3结构示意图
(a) Structure of Res2C3

(b) VANC3结构示意图
(b) Structure of VANC3

图 5　Res2C3 和 VANC3 结构示意图

Fig. 5　Structure of Res2C3 and VANC3

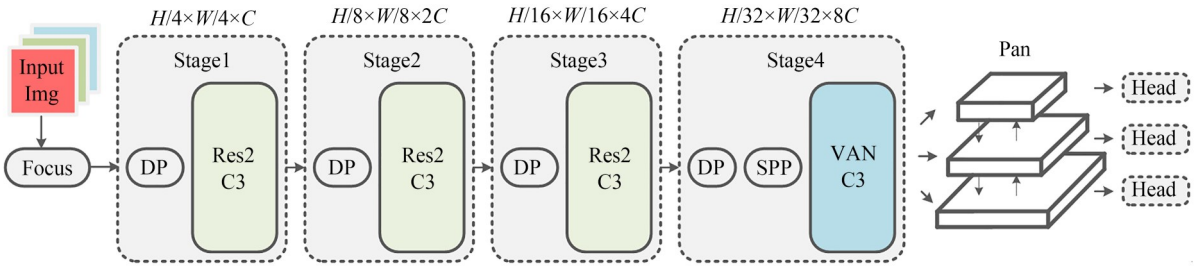


图 6　改进后的 Yolov5s 结构示意图

Fig. 6　Structure of improved Yolov5s

4　车辆检测实验

在实验中将原始 Yolov5s 和改进后的 Yolov5s 同 Yolov4-tiny 和 Yolov3-tiny 进行对比，上述两个网络与 Yolov5s 同样都是轻量级网络。除此以外还选择 Lin T Y[29]等人提出的经典单阶段网络 RetinaNet 参与对比测试，RetinaNet 广泛出现在近年各个目标检测任务的基准测试中。

4.1　实验环境

本文全部的实验均在表 2 配置和环境下运行。

表 2　实验配置以及实验环境

Tab. 2　Experimental configuration and environment

实验环境	具体配置信息
系统	Centos7
Python 版本	3.8
GPU	GTX 1080ti×2
Cuda 版本	10.1
实验框架	Pytorch 1.7.1

4.2　实验数据集

实验在两个常用的目标检测数据集上进行。

Pascal VOC 2012 数据集共约 1.7×10^4 张，共包含 20 个类别，实验提取数据集中小汽车、公交车、火车、自行车、摩托车共 5 个车辆类别，共计 3 073 张，按照 8∶2 比例随机划分训练集验证集。训练集共 2 459 张，验证集共 614 张。除此以外，从 Pascal VOC 2007 测试集上选取上述类别，并随机抽取 614 张图片作为测试集。

Microsoft COCO 2017 数据集共约 12×10^4 张，实验提取数据集中自行车、小汽车、摩托车、公交车、火车、卡车共 6 个车辆类别，并随机选取 7 000 张，按照 8∶1∶1 比例随机划分训练集、验证集、测试集。训练集 5 600 张，验证集测试集各 700 张。

4.3　实验配置

实验算法基于 Yolov5 5.0 官方仓库，在各项实验中均采用 640×640 图片大小进行训练，batchsize 为 32，Pascal VOC 车辆数据集共训练 400 轮，MS COCO 车辆数据集共训练 200 轮，优

化器为SGD，初始化学习率为1×10^{-2}，动量大小为0.937，权重衰减为5×10^{-4}。在实验中服从余弦cosine[30]学习率衰减，并且进行3轮学习率预热。默认启用马赛克增强，从头训练。1080ti单卡进行推理测试。mAP评判标准为测试集$IoU_{0.5}$。

Retinanet算法基于mmDetection平台，版本号2.15，训练图片大小为1 000×600，batchsize为4，初始化学习率为5×10^{-3}，动量大小为0.9，权重衰减为1×10^{-4}。在实验中服从余弦学习率衰减。其中Focal Loss超参数gamma值为2.0、alpha值为0.25。在训练阶段，对RetinaNet的特征提取网络进行权重初始化，默认采用由torchvision官方提供的ResNet50[31] ImageNet预训练权重，默认冻结第一阶段。网络其他层默认初始化为Xavier[32]。

4.4 实验结果

在后文所提及的全部实验表格中，Model Name为模型名称，Img Size为图片大小，mAP为各类别平均准确率均值，#Params为网络参数量，Flops为网络计算量，Inference time为推理一张图所消耗的时间。将Yolov5s的最后一个阶段改进为VANC3、前3个阶段改进为Res2C3，且特征图子集为4时命名为Yolov5s-improved。实验结果由表3所示。

表3 Pascal VOC和COCO车辆数据集各网络实验结果

Tab. 3 Experimental results of each networks on Puscal VOC vehicle dataset

Model name	Img size	#Params/M	FLOPs/G	mAP(%) Pascal VOC	mAP(%) MS COC	Inference time/ms
Yolov5s	640×640	7.3	17.1	82.5	62.3	6.1
Yolov5s-improved	640×640	8.1	17.3	84.6	64.0	6.9
Yolov4-tiny	640×640	6.4	17.4	70.3	54.3	3.9
Yolov3-tiny	640×640	8.9	13.3	70.2	51.5	3.7
Retinanet	1 000×600	36.2	122.2	84.4	61.2	52.8

由表3可以看出，在Pascal VOC车辆数据集上，Yolov5s-improved相较于原始Yolov5s提升2.1% mAP，mAP超越Yolov3-tiny和Yolov4-tiny约13%左右。与单阶段典型网络RetinaNet相比mAP小幅超越0.2%。

在MS COCO车辆数据集上，Yolov5s-improved相较于原始Yolov5s最高可提升1.7%mAP，mAP超越Yolov3-tiny和Yolov4-tiny约10%左右。RetinaNet的mAP甚至低于原始的Yolov5s，并且RetinaNet的参数量和计算量高达36.2M和122.2GFLOPs，分别约为改进后网络的4倍和7倍。除此以外，RetinaNet网络推理速度较慢，速度比改进前后的Yolov5s约慢7倍以上。

图7(a)和(b)分别为Pascal VOC车辆数据集改进前后PR曲线，(c)和(d)分别为MS COCO车辆数据集的PR曲线。每张PR曲线图(彩图见期刊电子版)的左下角为各类别的mAP，不同颜色的曲线对应不同类别。除了Pascal VOC车辆数据集下公交车类别，其余情况下两个数据集各个类别的mAP都可以得到不同程度的提升。各个类别的mAP可通过计算曲线与下方的面积获得，计算过程可表示为：

$$\mathrm{mAP}=\int_0^1 p(r)\mathrm{d}r. \tag{5}$$

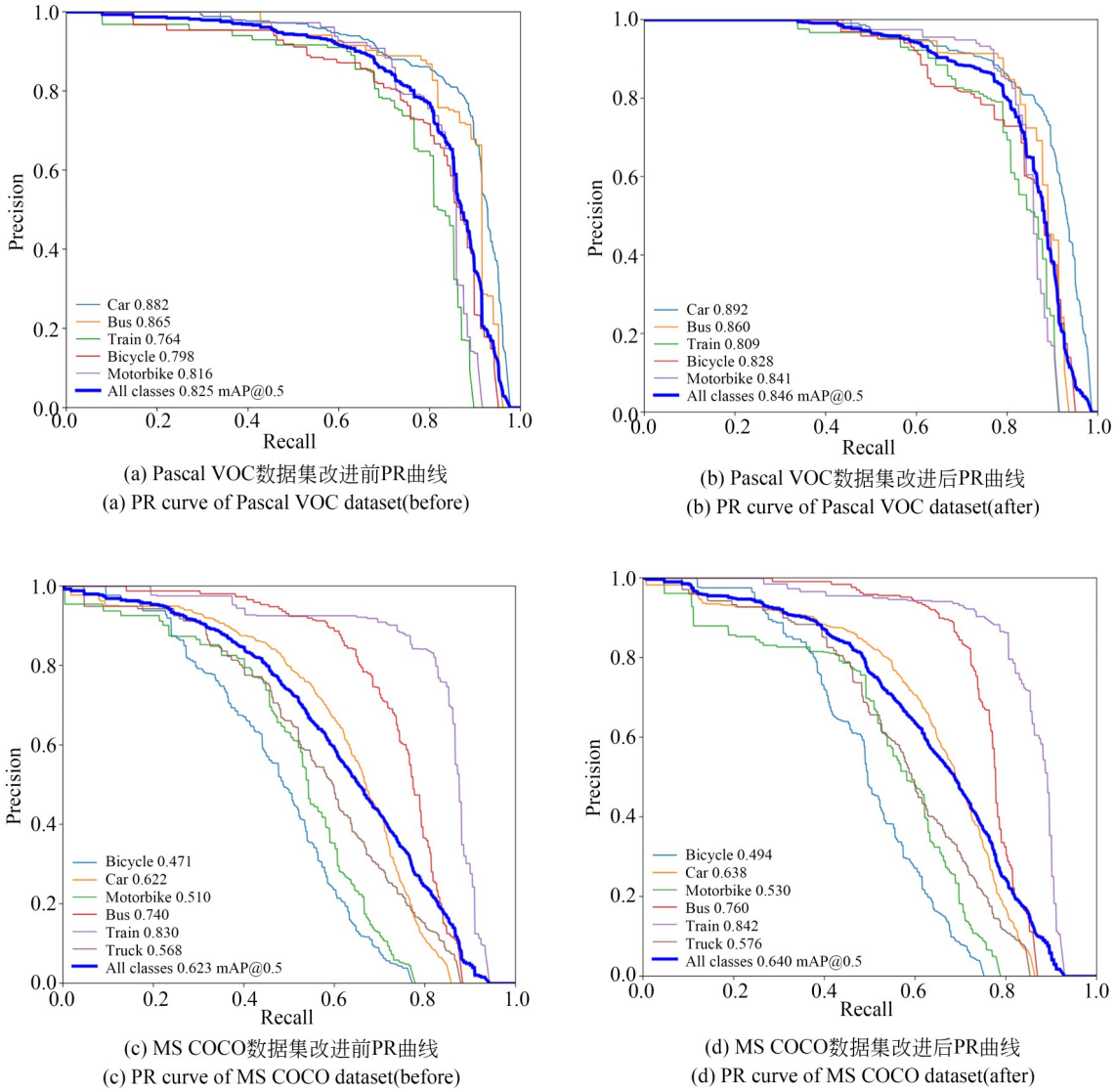
改进后Yolov5s深蓝色曲线与x轴围成的面积会比改进前更大，由此证明改进后算法是有效的。

4.5 消融实验

为探究两项改进泛化性，分别在两个数据集上进行消融实验。表4为Pascal VOC和MS COCO两个车辆数据集的消融实验。在Pascal VOC车辆数据集，VANC3可为网络带来0.7% mAP性能提升，网络计算量上涨0.6 GFLOPs。两项改进同时启用时，网络计算量比原网络上涨0.2 GFLOPs。在MS COCO车辆数据集，VANC3可为网络带来0.5% mAP性能提升。

4.6 更多实验

本节在Pascal VOC车辆数据集执行更多实验，用以尝试多种结构对网络性能影响。在网络



(a) Pascal VOC数据集改进前PR曲线
(a) PR curve of Pascal VOC dataset(before)

(b) Pascal VOC数据集改进后PR曲线
(b) PR curve of Pascal VOC dataset(after)

(c) MS COCO数据集改进前PR曲线
(c) PR curve of MS COCO dataset(before)

(d) MS COCO数据集改进后PR曲线
(d) PR curve of MS COCO dataset(after)

图 7　Pascal VOC 和 MS COCO 车辆数据集 PR 曲线

Fig. 7　PR curve of the Pascal VOC and MS COCO vehicle dataset

表 4　Pascal VOC 和 MS COCO 车辆数据集消融实验

Tab. 4　Ablation Experiments on Pascal VOC and MS COCO vehicle dataset

VANC3	Res2C3	FLOPs /G	mAP /% VOC	mAP /% COCO	Inference time/ms
×	×	17.1	82.5	62.3	6.1
√	×	17.7	83.2	62.8	6.2
√	√	17.3	84.6	64.0	6.9

第 1～4 阶段，依次将原有的 C3 改进为 VANC3，用以验证在不同位置的情况下 VANC3 对网络带来的影响。表 5 证明 VANC3 在高维优于其他三种情况。

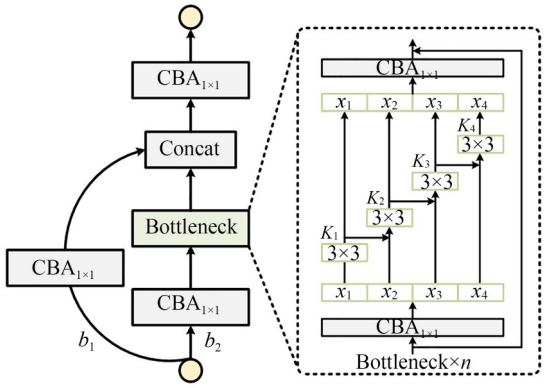
实验还尝试 Res2C3 残差模块不同分割方式或是增加卷积单元，用以探索 Res2C3 对网络性

表 5　VANC3 在不同阶段对 mAP 的影响

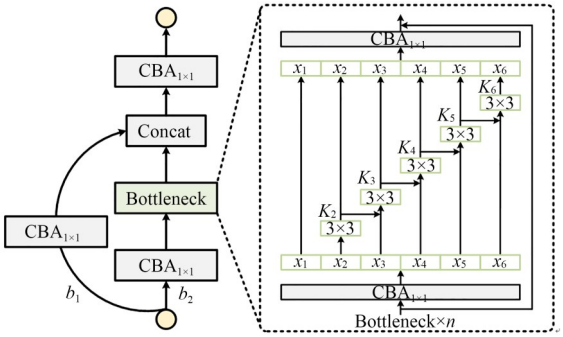
Tab. 5　Influence of VANC3 for mAP on different stages

Stage1	Stage2	Stage3	Stage4	mAP/%
√	×	×	×	82.8
×	√	×	×	82.0
×	×	√	×	81.7
×	×	×	√	83.2

能的影响。图 8(a)为在 Yolov5s-improved 的基础上，在x_1所在分支追加K_1卷积单元，图 8(b)为将 Yolov5s-improved 的 Res2C3 残差模块划分成$13w\times 6s$特征子集的结构示意图。



(a) 在K_1分支追加卷积单元
(a) Add a convolution units to the K_1 branch



(b) 将Res2C3划分成更多特征子集
(b) Split Res2C3 into more feature subsets

图 8　Res2C3 不同结构

Fig. 8　Different structure of Res2C3

表 6 为不同情况下 Res2C3 残差结构对网络各项性能的影响。追加K_1卷积单元后反而在 Pascal VOC 车辆数据集上 mAP 收益降低。采用$13w\times 6s$特征子集划分方式时发现网络性能并不如期待的那样会有更高的提升，反而网络的推理速度还会受到较大影响。

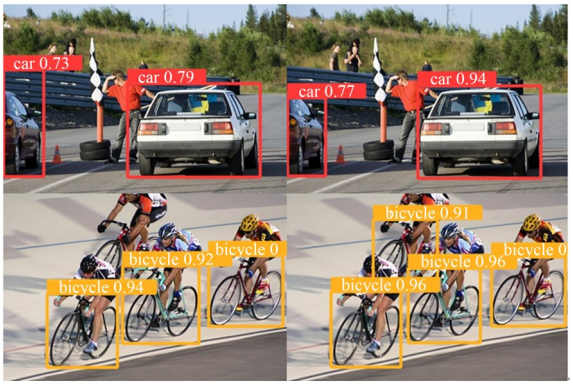
表 6　Res2C3 更多不同结构对 mAP 的影响

Tab. 6　Influence of Res2C3 residual module on mAP for more different structures

Model Name	Flops/G	mAP/%	Inference time/ms
org	17.1	82.5	6.1
add K_1	17.9	84.2	7.0
13w×6s	19.2	84.4	8.2

4.7　推理结果

图 9(a)和(b)推理结果分别来自 Pascal VOC 和 MS COCO 车辆测试集，左右分别为改进前后。



(a) Pascal VOC数据集推理结果
(a) Inference result on Pascal VOC dataset

(b) MS COCO数据集推理结果
(b) Inference result on MS COCO dataset

图 9　推理结果

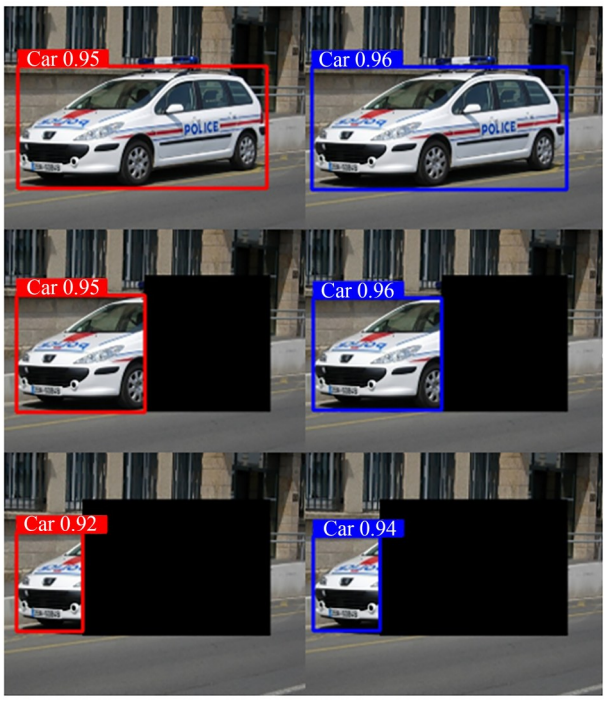
Fig. 9　Inference result

图 9(a)可以清晰的看出，改进后的网络在检测互相覆盖的 4 辆自行车时，可以很好地检测出图中间偏上位置的自行车，并可以提供更高置信度，而原网络并未能检测出自行车所在位置。图 9(b)可看出，网络在检测摩托车和黑色小汽车时，置信度更高。实验证明在两个不同的数据集上，网络的精度都可得到不同程度的提升，改进后的网络具有一定泛化性。

4.8　抗遮挡能力测试

本节专门对网络改进前后的抗遮挡能力进行测试。图 10(a)和(b)分别来自 Pascal VOC 和 MS COCO 车辆测试集，从上到下依次为原图、对车辆遮挡 50%、和对车辆遮挡 75%，每组左右两侧分别为原始 Yolov5s 和改进后 Yolov5s 的测试结果。

实验结果证明无论是在原始图片中，还是在遮挡 50% 和 75% 的情况下，改进后的 Yolov5s 都有着比原始网络更强的抗遮挡能力。



(a) Pascal VOC数据集抗遮挡效果测试
(a) Anti occlusion ability test on Pascal VOC dataset

(b) MS COCO数据集抗遮挡效果测试
(b) Anti occlusion ability test on MS COCO dataset

图 10　抗遮挡效果测试

Fig. 10　Anti occlusion ability test

5　结　论

本文提出一种基于 Yolov5s 的车辆检测改进算法，在高维时通过视觉注意力网络捕获长距离依赖，增强抗遮挡能力，并且在残差模块内部再次构造水平方向残差，增加网络的多尺度表达能力。改进后的网络在 Pascal VOC 车辆数据集上提升 2.1% mAP，在 MS COCO 车辆数据集上提升 1.7% mAP，网络整体依旧维持原有轻量化水平，与原网络相比具有更强的抗遮挡能力和更强的多尺度表达能力，且与其他轻量化网络相比检测效果显著，实现轻量化网络对车辆的高效检测。

参考文献：

[1] 孙敏，李免，赵玉舟，等. 基于实时交通状况和自适应像素分割的运动车辆检测[J]. 液晶与显示，2021，36(10)：1454-1462.
SUN M，LI M，ZHAO Y ZH，*et al*. Vehicle detection based on real-time traffic condition and adaptive pixel segmentation[J]. *Chinese Journal of Liquid Crystals and Displays*，2021，36(10)：1454-1462.(in Chinese)

[2] 张浩，杨坚华，花海洋. 基于FVOIRGAN-Detection的车辆检测[J]. 光学 精密工程，2022，30(12)：1478-1486.
ZHANG H，YANG J H，HUA H Y. Vehicle detection based on FVOIRGAN-Detection[J]. *Opt. Precision Eng.*，2022，30(12)：1478-1486.(in Chi-

nese)

[3] 杨慧. 基于HOG和Haar-like融合特征的车辆检测[D]. 南京：南京邮电大学，2013.
YANG H. *Vehicle Detection Based on HOG and Haar-Like Features*[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2013. (in Chinese)

[4] 李星，郭晓松，郭君斌. 基于HOG特征和SVM的前向车辆识别方法[J]. 计算机科学，2013，40(S2)：329-332.
LI X, GUO X S, GUO J B. HOG-feature and SVM based method for forward vehicle recognition [J]. *Computer Science*, 2013, 40(S2): 329-332. (in Chinese)

[5] 余永维，韩鑫，杜柳青. 基于Inception-SSD算法的零件识别[J]. 光学 精密工程，2020，28(8)：1799-1809.
YU Y W, HAN X, DU L Q. Target part recognition based Inception-SSD algorithm[J]. *Opt. Precision Eng.*, 2020, 28(8): 1799-1809. (in Chinese)

[6] 王宸，张秀峰，刘超，等. 改进YOLOv3的轮毂焊缝缺陷检测[J]. 光学 精密工程，2021，29(8)：1942-1954.
WANG CH, ZHANG X F, LIU CH, *et al*. Detection method of wheel hub weld defects based on the improved YOLOv3 [J]. *Opt. Precision Eng.*, 2021, 29(8): 1942-1954. (in Chinese)

[7] 李经宇，杨静，孔斌，等. 基于注意力机制的多尺度车辆行人检测算法[J]. 光学 精密工程，2021，29(6)：1448-1458.
LI J Y, YANG J, KONG B, *et al*. Multi-scale vehicle and pedestrian detection algorithm based on attention mechanism [J]. *Opt. Precision Eng.*, 2021, 29(6): 1448-1458. (in Chinese)

[8] 秦鸿睿. 基于改进EfficientDet的车辆检测算法研究[D]. 合肥：安徽建筑大学，2022.
QIN H R. Research on *Vehicle Detection Algorithm Based on Improved EfficientDet*[D]. Hefei: Anhui Jianzhu University, 2022. (in Chinese)

[9] 张洋. 改进的YOLOv3-tiny在城市交叉路口车辆检测中的应用[D]. 重庆：重庆师范大学，2020.
ZHANG Y. *Application of Improved YOLOv3-tiny in Vehicle Detection at Urban Intersections*[D]. Chongqing: Chongqing Normal University, 2020. (in Chinese)

[10] 张云强. 基于YOLOv4-tiny的车辆检测与测距算法研究[D]. 哈尔滨：哈尔滨理工大学，2022.
ZHANG Y Q. *Resarch on Detection and Distance Measurement Algorithm of Vehicles Based on YOLOv4-tiny*[D]. Harbin: Harbin University of Science and Technology, 2022. (in Chinese)

[11] 赵璐璐，王学营，张翼，等. 基于YOLOv5s融合SENet的车辆目标检测技术研究[J]. 图学学报，2022，43(5)：776-782.
ZHAO L L, WANG X Y, ZHANG Y, *et al*. Vehicle target detection based on YOLOv5s fusion SENet[J]. *Journal of Graphics*, 2022, 43(5): 776-782. (in Chinese)

[12] 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*). *June* 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 779-788.

[13] REDMON J, FARHADI A. YOLO9000: Better, Faster, Stronger[C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *July* 21-26, 2017, *Honolulu*, *HI*, *USA*. IEEE, 2017: 6517-6525.

[14] REDMON J, FARHADI A. YOLOv3: an incremental improvement [EB/OL]. 2018: arXiv: 1804.02767. https://arxiv.org/abs/1804.02767

[15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020: arXiv: 2004.10934. https://arxiv.org/abs/2004.10934

[16] 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*. *June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 8759-8768.

[17] WANG C Y, MARK LIAO H Y, WU Y H, *et al*. CSPNet: a New Backbone That Can Enhance Learning Capability of CNN [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (*CVPRW*). *June* 14-19, 2020, *Seattle*, *WA*, *USA*. IEEE, 2020: 1571-1580.

[18] HE K M, ZHANG X Y, REN S Q, *et al*. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(9): 1904-1916.

[19] TAN M X, LE Q V. EfficientNet: rethinking

model scaling for convolutional neural networks [EB/OL]. 2019: arXiv: 1905.11946. https://arxiv.org/abs/1905.11946

[20] TAN M X, LE Q V. EfficientNetV2: smaller models and faster training[EB/OL]. 2021: arXiv: 2104.00298. https://arxiv.org/abs/2104.00298

[21] VASWANI A, SHAZEER N, PARMAR N, *et al*. Attention is All You Need[C]. *Proceedings of the 31st International Conference on Neural Information Processing Systems. December* 4 - 9, 2017, *Long Beach*, *California*, *USA*. *New York*: *ACM*, 2017: 6000-6010.

[22] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, *et al*. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. 2020: arXiv: 2010.11929. https://arxiv.org/abs/2010.11929

[23] GUO M H, LU C Z, LIU Z N, *et al*. Visual attention network [EB/OL]. 2022: arXiv: 2202.09741. https://arxiv.org/abs/2202.09741

[24] ZHANG H, WU C R, ZHANG Z Y, *et al*. ResNeSt: split-attention networks [EB/OL]. 2020: arXiv: 2004.08955. https://arxiv.org/abs/2004.08955

[25] HU J, SHEN L, SUN G. Squeeze and Excitation Networks[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition. June* 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 7132-7141.

[26] LI D, YAO A B, CHEN Q F. PSConv: squeezing feature pyramid into one compact poly-scale convolutional layer[J]. *Computer Vision*, 2020: 615-632.

[27] DUTA I C, LIU L, ZHU F, *et al*. Pyramidal convolution: rethinking convolutional neural networks for visual recognition[EB/OL]. 2020: arXiv: 2006.11538. https://arxiv.org/abs/2006.11538

[28] GAO S H, CHENG M M, ZHAO K, *et al*. Res2Net: a new multi-scale backbone architecture [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(2): 652-662.

[29] LIN T Y, GOYAL P, GIRSHICK R, *et al*. Focal Loss for Dense Object Detection [C]. 2017 *IEEE International Conference on Computer Vision* (*ICCV*). *October* 22-29, 2017, *Venice*, *Italy*. IEEE, 2017: 2999-3007.

[30] LOSHCHILOV I, HUTTER F. SGDR: stochastic gradient descent with warm restarts[EB/OL]. 2016: arXiv: 1608.03983. https://arxiv.org/abs/1608.03983

[31] HE K M, ZHANG X Y, REN S Q, *et al*. Deep Residual Learning for Image Recognition [C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*). *June* 27-30, 2016, *Las Vegas*, *NV*, *USA*. IEEE, 2016: 770-778.

[32] TOLSTIKHIN I, HOULSBY N, KOLESNIKOV A, *et al*. MLP-mixer: an all-MLP architecture for vision [EB/OL]. 2021: arXiv: 2105.01601. https://arxiv.org/abs/2105.01601

作者简介：

荆修平(1996－)，男，辽宁鞍山人，硕士研究生，2018年于辽宁科技大学获得学士学位，主要从事机器视觉等方面研究。E-mail: jxp_snowman@163.com

通讯作者：

田　莹(1971－)，女，辽宁沈阳人，博士，教授，硕士生导师，2008年于沈阳工业大学获得博士学位，主要从事计算机视觉，数字图像处理，模式识别等方面研究。E-mail: astianying@126.com