引文格式：王笑,秦海超. 结合二阶池化注意与动态上采样的轻型街景车辆检测方法[J]. 地理空间信息, 2025, 23(9): 136-140. doi:10. 3969/j.issn. 1672-4623.2025.09.030

结合二阶池化注意与动态上采样的轻型街景车辆检测方法

王　笑[1]，秦海超[1]
（1. 建设综合勘察研究设计院有限公司，北京 100007）

摘　要：为实现低功耗硬件环境下的街景影像车辆检测，提出了一种轻量级检测方法。以MobileNetV3为特征提取网络，重构特征提取层并引入全局二阶池化卷积注意力单元；在MobileNetV3末端加入焦点调制单元聚合局部到全局尺度的上下文信息；利用逐点卷积与深度卷积核构建轻量级特征融合网络，并采用动态上采样单元进行图像上采样；在检测输出端采用质量焦点损失函数计算分类损失。伯克利深度驾驶100 K数据集的实验结果表明，该模型的平均精度均值达到89.64%；在测试环境下，检测速度达到37.13 F/s，比MobileNetV3提升了1.19 F/s，能在低功耗硬件环境下实现高效精准的街景影像车辆检测。

关键词：街景影像车辆检测；MobileNetV3；全局二阶池化卷积注意力；焦点调制；动态上采样；质量焦点损失函数
中图分类号：P231　　**文献标志码：**B　　**文章编号：**1672-4623（2025）09-0136-05

Lightweight Street View Vehicle Detection Method Combining Second Order Pooling Attention and Dynamic Upsampling

WANG Xiao[1], QIN Haichao[1]
(1. Comprehensive Institute of Geotechnical Investigation and Surveying Co., Ltd., Beijing 100007, China)

Abstract: To achieve vehicle detection in street view images under low-power hardware environment, we proposed a lightweight detection method. Taking MobileNetV3 as the feature extraction network, we reconstructed the feature extraction layer, and introduced global second order pooling（GSOP）convolutional attention units. At the end of MobileNetV3, we added a focus modulation unit to aggregate contextual information from local to global scales. We used point wise convolution and deep wise convolution kernel to build a lightweight feature fusion network, and used dynamic upsampling units for image upsampling. At the detection output, we used the quality focus loss function to calculate the classification loss. The experimental results on the BDD100K dataset show that the proposed model mAP@0.5 reaches 89.64%. In the testing environment, the detection speed reaches 37.13 F/s, which is 1.19 F/s faster than MobileNetV3, and can achieve efficient and accurate vehicle detection in street view images under low-power hardware environment.

Key words: vehicle detection in street view image, MobileNetV3, GSOP convolutional attention, focus modulation, dynamic upsampling, quality focus loss function

近年来以卷积神经网络（CNN）与目标检测转换器为代表的目标检测算法得到广泛应用[1-2]。CNN与车载高速相机采集的街景影像相结合能为道路目标感知、交通流量统计等场景提供有力的技术支持[3]。为实现高精度检测，当前主流目标检测算法大多通过深层次且庞大的模型挖掘目标特征[4]，推理过程会消耗大量计算资源[5]；而在车载嵌入式硬件等算力有限的部署环境，很难提供充足的双精度浮点算力。针对低功耗场景，部分学者[6-8]提出了以移动网络（MobileNet）为代表的轻量级目标检测网络，具有浅层紧凑的结构设计，消耗的算子较低。街景影像车辆检测任务对检测效率与精度都有较高要求，在车载相机视角下存在较多小尺度、高度重叠复杂的汽车样本，仅通过MobileNet等轻量级网络，难以充分捕获该类目标完整的上下文语义信息与细节特征。在特征上采样融合阶段，现有模型多通过线性插值或反卷积进行上采样，难以充分恢复目标细节信息；在检测输出端，现有模型大多采用静态损失函数，导致模型不能充分关注较难分类的样本，并存在训练与测试阶段结果不一致的情况。鉴于此，本文构建了一种轻量级街景影像车辆检测模型，以MobileNetV3为特征提取网络，重构特征提取层结构，并引入全局二阶池化卷积注意力（GSoP）单元；在特征提取网络末端增加焦点调制（FM）层，进一步融合局部与全局尺度的目标上下文

第一作者简介：王笑（1992—），本科，工程师，从事测绘工程测量、航空摄影测量等方面的工作，E-mail：Wx_serior@163.com。

信息。以路径聚合网络为框架，通过逐点卷积核（PWConv）、深度卷积核（DWConv）和动态上采样层（DwUpSample）搭建轻量级特征融合网络。在目标回归阶段，通过质量焦点损失（QFL）函数，使模型更偏向难分类样本的同时，训练与测试阶段的精度保持一致。

1　轻量级街景影像车辆检测方法

1.1　轻量级特征提取网络

MobileNetV3[2]采用神经架构搜索技术[3]设计，特征提取层由瓶颈结构深度可分离卷积核（DSC）构成，避免卷积核遍历特征图全部通道带来的冗余计算。在特征提取层末端，MobileNetV3通过压缩—激励（SE）单元捕获通道维度的注意力权重。SE单元在计算通道注意力时仅对特征图进行一阶全局平均池化，导致部分困难样本的特征无法被充分学习。为降低特征提取过程的运算消耗，本文重构了MobileNetV3的特征提取层，采用按通道分割操作，仅将半数通道参与特征提取。特征提取采用先升后降的逆向瓶颈结构，通过PWConv实现特征图通道的升降维，以高维度特征图实现特征提取，确保获得更丰富的目标特征；提取则由DWConv实现。完成特征提取后，通过通道拼接融合两组分割特征图，并借助通道混排操作打乱特征图通道顺序，保证后续操作中特征图内的信息交互。在重构基础上，引入GSoP单元通过二阶池化获得特征图协方差矩阵，捕获更深层次的通道权重关系。GSoP首先以1×1的常规卷积核对输入特征图进行通道降维；再采用二阶全局平均池化得到$c'\times c'$的协方差矩阵，矩阵中第i行元素表示第i个通道对其余通道的权重依赖，在特征激励阶段通过逐行卷积的方式从协方差矩阵中恢复特征图的结构信息；然后采用1×1的常规卷积核调整其通道，获得与各通道对应的注意力权重；最后通过逐通道相乘的方式，对输入特征图进行加权计算。为节约算力资源，本文仅在特征提取网络的第2~5层增加GSoP注意力单元。

为归一化任意尺度的输入特征图，同时融合局部到全局尺度的目标上下文信息，在特征提取网络末端，引入FM层。FM层通过焦点上下文化、门聚合（GA）、逐像素仿射变换3个步骤模拟视觉转换器在计算自注意力过程中的查询（Q）、键（K）、值（V）3组令牌交互过程。输入特征图首先通过线性计算映射到带有线性层的特征空间，再经过L层深度卷积来捕获由不同尺度的局部上下文信息，然后通过全局平均池化（GAP）来捕捉全局尺度的上下文信息，各层的计算公式为：

$$\begin{cases} Z^{l}=\mathrm{GeLU}(\mathrm{DWConv}(Z^{l-1})) \\ Z^{l+1}=\mathrm{GAP}(Z^{l}) \end{cases} \tag{1}$$

式中，$Z^{l},l\in[0,l]$为第l层输出特征，包含不同尺度的局部上下文信息；Z^{l+1}为全局平均池化后，带有全局上下文信息的特征图；GeLU为高斯误差线性单元激活函数。

捕获上下文信息后，利用GA操作聚集并筛选特征，首先通过线性层处理输入特征图（Z^{out}）来获取垂直水平方向的门控权重G，然后通过逐元素相乘的方式对Z^{l}赋权并将所有结果求和。在聚合上下文语义的基础上，通过线性计算层对Z^{out}进行处理，得到调制器$M=h(Z^{\mathrm{out}})$，最终通过仿射变换，将聚合的调制器注入到查询令牌中。输出特征图的计算公式为：

$$Y=Q(x_i)\otimes h\left(\sum_{l=1}^{L+1}g_i^{l}\otimes z_i^{l}\right) \tag{2}$$

式中，$Q(x_i)$为对令牌x_i的查询；h为线性变换；g_i^{l}与z_i^{l}分别为G^{l}和Z^{l}在i处的门控值与特征值。

改进的MobileNetV3结构见图1。

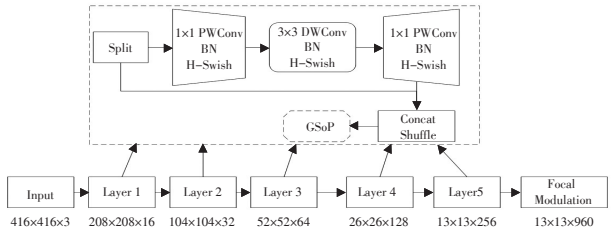


图1　轻量级特征提取网络结构图

1.2　轻量级特征融合网络

路径聚合网络（PANet）将来自特征提取网络后3层的特征图经过自底向上与自顶向下的双重采样融合得到不同层级的特征图。在上采样过程中，PANet通过双重线性插值算法重建大尺寸特征图，但双线性插值不能根据特征图内信息变化实时学习特征，导致重建特征图内的信息不够精确[9]，因此本文以PANet为基础，重构了轻量级特征融合网络。重构特征融合网络采用1×1的PW卷积核进行特征图通道处理，3×3的DW卷积核进行特征图下采样，以降低特征图处理的算力消耗。为实现特征精细化重建，引入动态上采样（DyU）单元，首先通过尺寸为$H\times W\times2gs^{2}$的线性变

换操作（Linear）对输入尺寸为 $H_1\times W_1\times C$ 的特征图进行通道压缩；再通过像素混洗层（PS）操作对压缩后的特征图进行偏移，得到 $sH_1\times sW_1\times 2g$ 的偏移矩阵 $\boldsymbol{O}$；然后按像素位置，与原始采样网格 $\boldsymbol{G}$ 逐点相加得到采样集 $\boldsymbol{S}$；最后以采样集 $\boldsymbol{S}$ 为模板，通过网格采样（Grid Sample）将输入特征图上采样至 $H_2\times W_2\times C$ 大小。在自顶向下的特征图采样路径中，PANet通常采用常规卷积核进行特征图下采样，为减少计算消耗，本文以DSC替代常规卷积核，最终输出3个不同尺寸的特征图，用于多尺度目标检测。

1.3 优化目标分类损失函数

本文采用YOLOv8模型的损失函数结构，损失部分由目标框回归损失、目标分类损失、目标置信度损失组成。由于在YOLOv8训练阶段，目标回归、分类、置信度3组函数均采用独立计算，但在推理阶段，3组损失函数将采用加权相乘的方式得到一个最终得分，这就导致训练与测试阶段的精度得分不一致[10]；此外，目标分类采用经典的焦点损失函数，由于目标框回归与目标分类任务的不一致[11]，导致部分高度重叠、小尺寸的复杂样本难以被检出。因此，本文引入QFL函数，并将分类得分由传统的(0,1)独热分布表示拓展为(0,1)区间内的连续值，即目标分类直接得分与目标回归得分的联合表示，以保持训练—测试阶段精度的一致性。通过动态调整损失函数，使模型在训练阶段更关注较难分类或定位的样本，提高正确检出率。QFL函数的计算公式为：

$$\mathrm{QFL}(\sigma)=-|y-\sigma|^{\beta}((1-y)\log(1-\sigma)+y\log\sigma) \quad (3)$$

式中，y 为目标类别真值，$y=0$ 表示分类得分为0的负样本，$0<y\leqslant 1$ 表示交并比（IoU）得分为 y 的真样本；σ 为对 y 进行S型曲线函数激活后的输出值；$\beta>0$，为控制平滑率的动态调节参数。

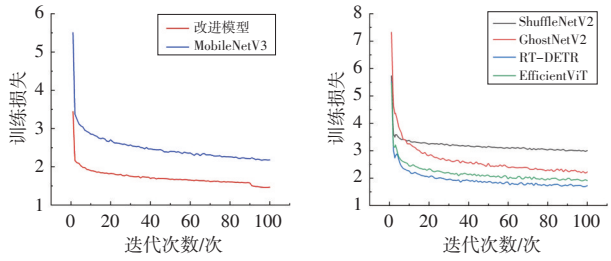
2 实验与结果分析

2.1 实验数据集

本文采用伯克利深度驾驶100K数据集（BDD100K）[12]来完成模型的训练与验证。BDD100K数据集包含不同街道场景的高分辨率图像，其中训练样本68 943张、验证样本9 882张、测试样本10 000张。

2.2 模型训练参数

为充分验证轻量化模型在小型硬件上的实际表现，采用“训练—部署分离”的策略，训练服务器的操作系统为Ubuntu20.04，CPU为Intel Xeon Platinum 8352V，GPU为Nvidia GeForce RTX 4090，服务器内存为64 GB，硬盘容量为1 TB。改进模型与对照组模型第二代混排网络（ShuffleNetV2）[13]、第二代幽灵网络（GhostNetV2）[14]、MobileNetV3、高效视觉转换器（EfficientViT）[15]、实时转换器目标检测模型（RT-DETR）[16]均在相同环境下完成训练，所有模型采用相同的训练超参数。训练过程中，输入图像尺寸统一为640×640，单批次送入网络的图像为16张，初始学习率为0.01，模型迭代训练100次，期间采用自适应矩估计优化算法优化训练参数，一阶动量参数为0.916，各模型的训练损失见图2，可以看出，所有模型均在训练过程中得到了充分收敛，但改进模型的初始损失更小、收敛速度更快，在训练结束阶段损失值更小。训练完成后，将所有模型封装转译，部署在带有ARM架构CPU的设备Nvidia Jetson Xavier上，开展性能测试实验。



a 改进模型与MobileNetV3　　b 对照组模型

图2 各模型训练损失曲线

2.3 评价指标

本文通过检测精度、检测速度、模型大小3个方面的指标对实验结果进行定量评价。本文选取精确率、召回率、不同IoU阈值下的精度均值（mAP）作为检测精度的评价指标；选取每秒检测帧数（FPS）作为检测速度的评价指标；选取模型运行参数量（Params）作为模型大小的评价指标。

2.4 实验结果分析

2.4.1 模型对比

为验证改进模型的综合性能，将其与对照组模型在BDD100K数据集上完成综合性能测试，结果见表1，可以看出，改进模型的mAP@0.5为89.64%，比

表1 对比实验结果

模型	mAP@0.5/%	mAP@0.5:0.95/%	FPS/F·s^{-1}	Params/M
ShuffleNetV2	76.28	63.50	32.15	8.46
GhostNetV2	79.74	68.71	42.45	6.58
MobileNetV3	82.57	71.42	35.94	7.41
EfficientViT	83.28	72.56	29.74	14.83
RT-DETR	87.31	74.75	19.74	45.42
改进模型	89.64	75.94	37.13	7.25

MobileNetV3模型提高了7.07%，比EfficientNet、ShuffleNetV2、GhostNetV2模型分别提高了21.17%、13.36%、9.9%，说明了改进模型在检测精度方面具备明显优势；在测试硬件环境下，改进模型的FPS达到了37.13 F/s，比MobileNetV3模型提高了3.31%，略低于GhostNetV2模型；改进模型的Param仅为7.25 M，比MobileNetV3模型减少了2.16%；综合上述结果，改进模型在街景影像车辆目标检测任务中实现了较高精度与较快速度的平衡，综合性能较理想。

2.4.2　消融实验

为进一步验证不同改进方法对原模型性能的影响，本文开展消融实验。以MobileNetV3模型为基准模型（Base），在其上重构GSoP、增加FM、DwUpSample、QFL四项改进措施，构建4组消融模型（Model1~4），Model5~7则逐步累加上述改进措施，Model7为完整改进模型。所有模型均在BDD100K的测试数据集上完成测试，结果见表2，Base在测试集上的mAP@0.5达到了82.57%；Model 1的mAP@0.5上升到85.58%，精确率、召回率均有提升；Model2~4的mAP@0.5较Base分别提高了2.18%、1.26%、1.74%；Model5~7的mAP@0.5%也在同步上升，Model7的mAP@0.5达到了89.64%，较Base提高了7.07%m；结果表明，单独或联合采用4种改进方法均能有效提高检测精度，其中重构GSoP带来的精度提升最明显（图3）。

表2　消融实验结果

模型	方法				精确率/%	召回率/%	mAP@0.5/%
	GSoP	FM	DwUpSample	QFL			
Base	–	–	–	–	81.46	77.25	82.57
Model 1	√	–	–	–	88.74	79.32	85.58
Model 2	–	√	–	–	86.43	78.37	84.75
Model 3	–	–	√	–	86.74	78.45	83.83
Model 4	–	–	–	√	81.57	77.69	84.31
Model 5	√	√	–	–	86.51	79.23	87.47
Model 6	√	√	√	–	89.48	80.54	88.32
Model 7	√	√	√	√	91.42	82.47	89.64



图3　不同消融模型检测结果对比

2.4.3　检测结果对比

对比改进模型与MobileNetV3在相同场景下的检测结果，以直观说明改进后模型性能的提升。由图4可知，对于复杂环境下尺寸较小或遮挡较多的物体，MobileNetV3模型存在较多的漏检，也会将部分无关背景中的要素误检为车辆样本；改进模型则对于弱光环境下的小尺寸目标和严重遮挡情况下的困难目标具有更好的检测能力。



图4　改进模型与MobileNetV3检测结果对比

3　结　语

为实现轻量化硬件条件下的街景影像车辆目标检测，本文提出了一种基于MobileNetV3的改进模型。重构MobileNetV3的特征提取层，并引入GSoP层，通过计算二阶全局平均池化得到特征图的协方差矩阵，进而计算各通道间更深层次的相互依赖关系。在特征提取网络末端，引入FM层捕获并融合局部到全局尺度的目标上下文信息。结合DWConv、PWConv与DyU单元，重构轻量级特征融合网络，在进一步降低特征融合算力消耗的同时，采用DyU单元进行更高效的轻量上采样。在目标分类过程中，采用QFL损失将IoU得分融入分类得分计算过程，使模型在训练与测试阶段保持精度的一致性。在公开街景影像数据集BDD100K上开展实验，结果表明改进模型的mAP@0.5较MobileNetV3模型提高了7.07%，且优于目前主流的轻量级目标检测模型；在测试环境下，改进模型的检测速度达到了37.13 F/s，较MobileNetV3模型同样有所提升。后续研究将引入红外波段等的街景图像拓展数据集，以提高模型的全天候检测能力；同时加入单目视觉算法，以实现场景中目标的检测与定位。

参考文献

[1]　余添添,吴松,唐芝青,等.利用铁塔视频图像和改进YOLOv5

的违规施工监测[J]. 地理空间信息,2024,22(4):45-48

[2] 黄少雄,兰建雄,谭勇,等. 基于无人机视频的公路桥梁边坡异常检测[J]. 地理空间信息,2024,22(1):78-81

[3] 侯鑫,王艳,王绚,等. 全景影像在城市研究中的应用进展综述[J]. 计算机科学与探索,2024,18(7):1 661-1 682

[4] 徐小华,周长兵,胡忠旭,等. 轻量级深度神经网络模型适配边缘智能研究综述[J]. 计算机科学,2024,51(7):257-271

[5] Zhang D, Hao X, Wang D, et al. An Efficient Lightweight Convolutional Neural Network for Industrial Surface Defect Detection[J]. Artificial Intelligence Review , 2023 , 56(9): 10 651-10 677

[6] 胡焱,赵宇航,胡皓冰,等. 基于YOLOv5s的轻量化红外图像行人目标检测研究[J]. 激光与红外,2024,54(2):295-301

[7] Zhang J, Cai Y Y, Yang D, et al. Mobilenetv3-BLS: A Broad Learning Approach for Automatic Concrete Surface Crack Detection[J]. Construction and Building Materials, 2023,392:131 941

[8] 林冲,范加利,闫文君,等. 深度神经网络架构轻量化方法综述[J]. 中国电子科学研究院学报,2024,19(2):179-193

[9] Yu H, Li X, Feng Y, et al. Multiple Attentional Path Aggregation Network for Marine Object Detection[J]. Applied Intelligence,2023,53(2):2 434-2 451

[10] Terven J, Córdova-Esparza D M, Romero-González J A. A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-nas[J]. Machine Learning and Knowledge Extraction , 2023 , 5(4): 1 680-1 716

[11] Liang X, Cheng W, Zhang C, et al. YOLOD: A Task Decoupled Network Based on YOLOv5[J]. IEEE Transactions on Consumer Electronics,2023,69(4):775-785

[12] Liu G, Han J. Cross-domain Object Detection Using Minimized Instance Shift Image-image Translation[J]. Visual Computer,2023,39(10):5 013-5 026

[13] Han J, Yang Y. L-Net: Lightweight and Fast Object Detector-based ShuffleNetV2[J]. Journal of Real-time Image Processing,2021,18(6):2 527-2 538

[14] Tang Y, Han K, Guo J, et al. GhostNetv2: Enhance Cheap Operation with Long-range Attention[J]. Advances in Neural Information Processing Systems,2022,35:9 969-9 982

[15] Liu X, Peng H, Zheng N, et al. Efficientvit: Memory Efficient Vision Transformer with Cascaded Group Attention[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2023:14 420-14 430

[16] Peng H, Chen S. FedsNet: The Real-time Network for Pedestrian Detection Based on RT-DETR[J]. Journal of Real-time Image Processing,2024,21(4):1-12

（上接第135页）

智能城市,2020,6(22):69-70

[6] 孟蕾,林超. 机载LiDAR技术生成DEM的质量检查与解决方案探讨[J]. 国土资源遥感,2020,32(1):7-12

[7] 肖春蕾,郭兆成,郑雄伟,等. 机载LiDAR技术在地质调查领域中的几个典型应用[J]. 国土资源遥感,2016,28(1):136-143

[8] 曹岳飞. 基于平面约束模型的机载激光测深雷达安置误差标定技术研究[D]. 青岛:山东科技大学,2018

[9] 张文旭. 无人机LiDAR系统误差检验检校技术研究[D]. 鞍山:辽宁科技大学,2022

[10] 周阳林,李广云,王力,等. 一种GNSS/INS/LiDAR组合导航传感器安置关系快速标定方法[J]. 中国惯性技术学报,2018,26(4):458-463

[11] 肖天中,张继宏,林金朝,等. 无人机载激光扫描系统标定方法与应用[J]. 应用激光,2019,39(2):328-332

[12] 亢甲杰,张福民,曲兴华. 激光雷达坐标测量系统的测角误差分析[J]. 激光技术,2016,40(6):834-839

[13] 马浩. 机载激光雷达测量系统误差检校方法研究[D]. 青岛:山东科技大学,2011

[14] 马广迪,刘正军,段敏燕. 基于安置角误差改正的机载激光雷达数据自检校方法研究[C]. 第一届全国激光雷达对地观测高级学术研讨会,2010

[15] 张靖,江万寿. 基于虚拟连接点模型的机载LiDAR系统安置误差自检校[J]. 测绘学报,2011,40(6):762-769

[16] 刘田龙,常虹,谢劭峰,等. 国产机载LiDAR系统集成误差检校方法的比较[J]. 地理空间信息,2015,13(2):28-30

[17] 王亮,刘鹏,贾旭斌. 一种利用点云数据提取建筑物屋顶面的方法[J]. 地理空间信息,2023,21(12):86-88

[18] 张小红. 机载激光雷达测量技术理论与方法[M]. 武汉:武汉大学出版社,2007

[19] 李世明. 机载LiDAR系统检校与应用探索分析[D]. 昆明:昆明理工大学,2013

[20] 郭玉芳,陈浩,巨小文,等. 机载激光雷达的标准化进展分析[J]. 地理空间信息,2024,22(2):13-16