ISSN 1003-501X CN 51-1346/O4

光电工程

Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊
Scopus CSCD

渐进式多粒度ResNet车型识别网络

徐胜军，荆扬，李海涛，段中兴，刘福友，李明海

引用本文:

徐胜军，荆扬，李海涛，等. 渐进式多粒度ResNet车型识别网络[J]. 光电工程，2023， **50**(7): 230052.

Xu S J, Jing Y, Li H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. *Opto-Electron Eng*, 2023, **50**(7): 230052.

https://doi.org/10.12086/oee.2023.230052

收稿日期: 2023-03-05; 修改日期: 2023-05-23; 录用日期: 2023-06-05

相关论文

多尺度注意力与领域自适应的小样本图像识别

陈龙，张建林，彭昊，李美惠，徐智勇，魏宇星

光电工程 2023, **50**(4): 220232 doi: 10.12086/oee.2023.220232

基于语义分割的实时车道线检测方法

张冲，黄影平，郭志阳，杨静怡

光电工程 2022, **49**(5): 210378 doi: 10.12086/oee.2022.210378

结合层次化搜索与视觉残差网络的光学舰船目标检测方法

徐安林，杜丹，王海红，张强，李雅哲

光电工程 2021, **48**(4): 200249 doi: 10.12086/oee.2021.200249

深度双重注意力的生成与判别联合学习的行人重识别

张晓艳，张宝华，吕晓琪，谷宇，王月明，刘新，任彦，李建军

光电工程 2021, **48**(5): 200388 doi: 10.12086/oee.2021.200388

更多相关论文见光电期刊集群网站

OEE 光电工程 Opto-Electronic Engineering

http://cn.oejournal.org/oee

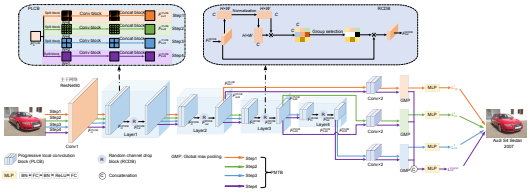
OE_Journal Website

Opto-Electronic Engineering
光 电 工 程

Article
2023 年, 第 50 卷, 第 7 期

DOI: 10.12086/oee.2023.230052

渐进式多粒度 ResNet 车型识别网络



徐胜军[1,2]，荆　扬[1,2*]，李海涛[3]，
段中兴[1,2]，刘福友[4]，李明海[1,2]

[1]西安建筑科技大学信息与控制工程学院，陕西 西安 710055；
[2]西安市建筑制造智能化技术重点实验室，陕西 西安 710055；
[3]江苏省交通工程建设局，江苏 南京 210004；
[4]中交隧道工程局有限公司，北京 100024

摘要：针对车辆因姿态、视角等成像差异造成车型难以识别问题，提出一种基于渐进式多粒度 ResNet 车型识别网络。首先，以 ResNet 网络作为主干网络，提出渐进式多粒度局部卷积模块，对不同粒度级别的车辆图像进行局部卷积操作，使网络重构时能够关注到不同粒度级别的车辆局部特征；其次，对多粒度局部特征图利用随机通道丢弃模块进行随机通道丢弃，抑制网络对车辆显著性区域特征的注意力，提高非显著性特征的关注度；最后，提出一种渐进式多粒度训练模块，在每个训练步骤中增加分类损失，引导网络提取更具辨别力和多样性的车辆多尺度特征。实验结果表明，在 Stanford cars 数据集、Compcars 网络数据集和真实场景下的车型数据集 VMRURS 上，所提网络的识别准确率分别达到了 95.7%、98.8% 和 97.4%，和对比网络相比，所提网络不仅具有较高的识别准确率，而且具有更好的鲁棒性。

关键词：车型识别；ResNet 网络；渐进式多粒度局部卷积；随机通道丢弃；渐进式多粒度训练

中图分类号：TP391.4　　**文献标志码：**A

徐胜军，荆扬，李海涛，等. 渐进式多粒度 ResNet 车型识别网络 [J]. 光电工程，2023，**50**(7): 230052
Xu S J, Jing Y, Li H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. *Opto-Electron Eng*, 2023, **50**(7): 230052

Progressive multi-granularity ResNet vehicle recognition network

Xu Shengjun[1,2], Jing Yang[1,2*], Li Haitao[3], Duan Zhongxing[1,2], Liu Fuyou[4], Li Minghai[1,2]

[1]College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, Shannxi 710055, China;
[2]Xi'an Key Labratory of Building Manufactaring Intelligent & Automation Technology, Xi'an, Shannxi 710055, China;
[3]Traffic Engineering Construction Bureau of Jiangsu Province, Nanjing, Jiangsu 210024, China;
[4]CCCC Tunel Engineering Company Limited, Beijing 100024, China

收稿日期：2023-03-05；修回日期：2023-05-23；录用日期：2023-06-05
基金项目：国家自然科学基金资助项目 (51678470, 61803293)；陕西省教育厅专项科研项目资助 (18JK0477, 2017JM6106)；陕西省自然科学基础研究计划资助项目 (2020JM-472, 2020JM-473, 2019JQ-760)；西安建筑科技大学基础研究基础资助项目 (JC1703, JC1706)；陕西省科技厅社发攻关项目 (2021SF-429)
*通信作者：荆扬，jingyang0525@xauat.edu.cn。

Abstract: Aiming at the problem that vehicle models are difficult to recognize due to differences in vehicle posture and viewing angles, a vehicle model recognition network based on progressive multi-granularity ResNet is proposed. Firstly, a progressive multi-granularity local convolution module is proposed by using the ResNet network as the backbone network to perform local convolution operations on vehicle images of different granularity levels, so that local features of vehicles at different granularity levels can be paid attention to when the network is reconstructed. Secondly, for the multi-granularity local feature map, the random channel discarding module is adopted to perform random channel discarding, which suppresses the network's attention to the vehicle's salient regional features and improves the attention of non-salient features. Finally, a progressive multi-granularity training module is proposed. A classification loss is added in each training step to guide the network to extract more discriminative and diverse vehicle multi-scale features. Experimental results show that the recognition accuracy of the proposed network reaches 95.7%, 98.8%, and 97.4% respectively on the Stanford-cars dataset, the Compcars network dataset, and the vehicle model dataset VMRURS in real scenes. In comparison with the comparative network, the proposed network not only has higher recognition accuracy but also has better robustness.

Keywords: vehicle model recognition; ResNet network; progressive multi-granularity local convolution block; random channel drop block; progressive multi-granularity training

1 引　言

车型识别旨在识别车辆的品牌、型号、年份等具体信息，能辅助验证跟踪车辆信息的准确性，因此被广泛应用于智能交通领域的车辆重识别、车辆跟踪等场景中。虽然近些年来车型识别研究已经取得了阶段性成果，但是由于车辆拍摄图像受其所处环境、天气、光照等影响干扰大，显著增加了车型识别的难度。同时，由于车型种类繁多，而且同一品牌不同型号的车型之间外观差异小，这为车型的准确识别带来很大挑战。

随着车型识别技术的发展，研究人员对车型识别任务进行了深入研究，提出了许多车型识别研究方法。主要分为基于传统机器学习的车型识别方法[1-6]和基于深度学习的车型识别方法[7-24]。基于传统车型图像识别的方法常利用浅层特征向量和分类器结合来解决车型识别问题。Liao 等[5]提出了一种基于车辆部件的分类方法，采用强监督 DPM 引入语义层次结构对车辆图像进行语义分割，基于分割部件的外观和语义来识别车辆。Hsieh 等[6]通过对感兴趣区域进行网格划分，对每个网格使用 HOG 和对称 SURF 描述算子提取特征，并在每个网格块上使用支持向量机 (SVM) 训练弱分类器进行车型识别。虽然上述方法取得了一定的效果，但是基于人工设计的方法存在特征难以描述、鲁棒性较差的问题，因此难以准确获得车辆更具判别力的细粒度关键特征，无法满足当前实际场景下的车型识别问题。

近些年来，由于计算机视觉在图像处理领域的飞速发展，更多的研究人员将目光转移到使用深度学习的方法来进行车型识别[7-9]。在深度学习任务中，车型识别具有类间差异小和类内差异大的特点，因此其识别任务属于典型的图像细粒度识别问题[10-11]。基于深度学习的车型识别方法主要分为：基于强监督学习的识别方法[12-15]和基于弱监督学习的识别方法[16-24]。基于强监督学习的识别方法是指利用数据集中所给出的标注信息来对测试集中图片的特征点进行定位，再对定位到的特征区域进行处理，进而得到最终的分类结果。Fang 等[12]提出了一种将粗粒度和细粒度特征结合的网络 (CFNet)，通过车辆关键性区域的定位，进而利用全局信息和局部特征进行车型识别。Zhang 等[13]提出了一种强监督语义对齐网络 (FOAT)，通过利用一种有效的零件对齐方式将语义先验结合到几何对齐中，进而实现车型识别。上述方法虽然取得了不错的识别效果，但由于强监督学习需要大量的人工标注样本，并且人工标注的位置具有很强的主观性，不一定是最佳判别区域，因此基于强监督学习的方法难以有效解决真实场景下的车型识别问题。

基于弱监督学习的识别方法常利用注意力机制[16-18]、双线性卷积神经网络[19-22]、度量学习[23-24]等方法来定位图像关键区域。相较于基于强监督学习的识别方法，基于弱监督学习的识别方法无需大量标注信息，仅利用图像标签即可取得较高的识别准确率，因此近年来受到了研究者的广泛关注。基于注意力机

制的方法通过引入注意力模块即可引导网络获取车辆的显著性特征。Ding 等[16] 提出了注意力金字塔卷积神经网络 (AP-CNN)，利用双路径层次结构使网络可以同时学习到车辆的低级细节特征和高级语义特征。Rao 等[17] 提出基于反事实注意力的识别网络 (CAL)，通过因果推理的手段来使网络学习到更加有效的细粒度特征。基于双线性卷积神经网络的方法通过外积相乘运算，利用协方差矩阵的特征表示与深度描述符相结合实现不同特征的融合以提升细粒度识别性能。Lin 等[19] 提出了双线性卷积神经网络 (bilinear CNN)，采用外积的方式对双通道特征进行融合，从而建模不同通道间的线性关系。Yu 等[20] 在 Bilinear CNN 的基础上提出了跨层双线性池化方法 (HBP)，该网络利用不同卷积层之间的层间特征交互，有效捕捉细粒度子类别之间的部分鉴别属性。基于度量学习的方法通过将数据点从原始的向量空间映射到一个新的向量空间，旨在拉近相似点之间的距离，拉远非相似点之间的距离。Sun 等[23] 提出了多注意力多类约束网络 (MAMC)，利用度量学习框架中的多注意力多类约束策略，将相同类别的相同注意力拉近，将不同注意力和不同类别拉远，提升网络对细粒度特征的辨别力。由于车辆特征信息具有多粒度特性，特征局部区域面积相差较大，如较大粒度的栅格、车轮胎，以及较小粒度的车标、门把手等，这些不同粒度的车辆特征对于车型识别问题来说是具有较高辨别力的关键特征，但是车辆不同粒度特征显著性差别较大，因此需要合理地将多粒度特征有效结合起来。

上述基于弱监督学习的分类方法大多采用注意力机制、双线性卷积神经网络、度量学习等策略，这些策略导致网络更多关注到了车辆的栅格、车轮胎等较大粒度的显著性判别区域，忽视了车标、车门把手等有辨别力的小粒度车辆特征，导致车辆因姿态、视角等成像差异造成车型难以识别问题。为解决上述问题，提出一种新的基于弱监督学习策略的渐进式多粒度 ResNet 网络车型识别方法。主要贡献如下:

1) 提出一种渐进式多粒度局部卷积模块，通过将网络进行不同粒度切分，迫使网络关注到不同粒度级别的车辆局部特征。

2) 提出一种随机通道丢弃模块，可以有效抑制网络聚焦到显著性区域的能力，通过将注意力分散，迫使网络关注车辆的非显著性区域。

3) 提出一种渐进式多粒度训练模块，通过将网络训练过程分为不同阶段，使得网络可以有效地整合提取到的车辆多粒度特征。

2 渐进式多粒度 ResNet 车型识别网络

为了利用更具辨别力的车辆多粒度特征信息提高车型识别准确率，提出一种基于渐进式多粒度 ResNet 车型识别网络，总体结构如图 1 所示。所提网络主要由 ResNet 主干网络、渐进式多粒度局部卷积模块 (progressive multi-granularity local convolution block, PLCB)、随机通道丢弃模块 (random channel drop block, RCDB)、渐进式多粒度训练模块 (progressive multi-granularity training block, PMTB) 四个部分组成。ResNet 主干网络用于捕获车辆图像特

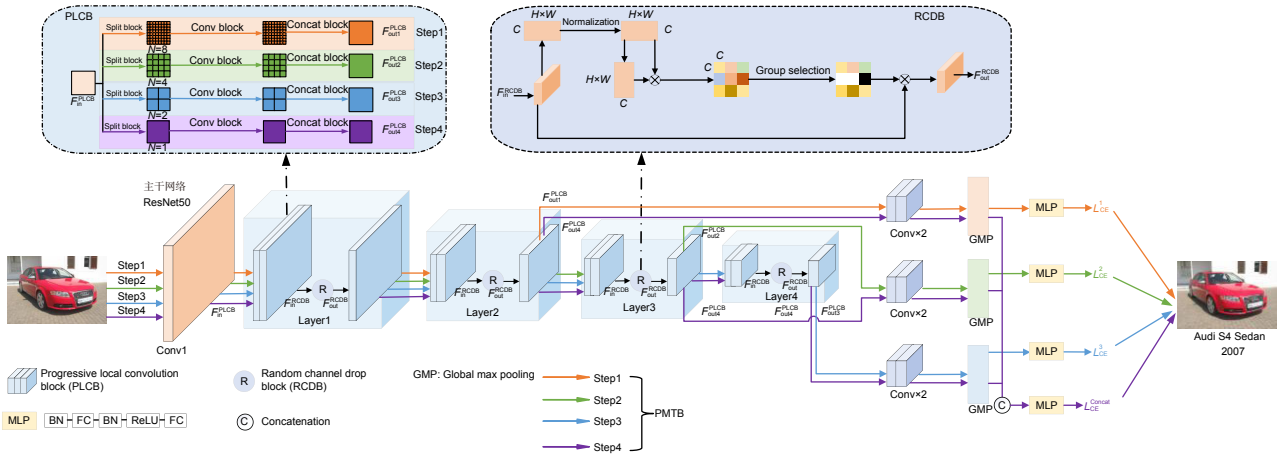


图 1 网络整体结构

Fig. 1 Overall structure of the proposed network

征；PLCB 设置在 ResNet 四层残差结构前，通过将车辆图像切分成不同大小的局部图像捕获车辆多粒度特征；RCDB 用于抑制网络对车辆显著性区域特征的注意力，提高非显著性特征的关注度；PMTB 将训练过程分为不同阶段，通过对不同阶段提取到的特征施加分类损失，引导网络捕获辨别性和多样性特征。具体地，图中 Conv1 表示 ResNet50 的第 1 层卷积层，PLCB 模块由 3 个步骤构成，分别为 Split block (n)，Conv block，Concat block (n)，其中 n 代表网格切分的大小，在不同阶段 n 的取值不同，Conv block 选用 ResNet50 的 4 层残差结构 Conv2~Conv5，Layer1~Layer4 代表主干网络提取特征的 4 个阶段，Conv*2 表示 2 个卷积层，分别为 1 个 1×1 卷积用于降维和 1 个 3×3 卷积用于调整通道数，经过 Conv*2 后通道数统一调整为 1024，GMP (global max pooling) 代表全局最大池化，MLP (multilayer perceptron) 代表多层感知机，由 2 个批量归一化 (batch normalization, BN)、1 个 ReLU 激活函数和 2 个全连接层 (fully connected layers, FC) 层构成。L_{CE}^{1}、L_{CE}^{2}、L_{CE}^{3}、L_{CE}^{Concat} 分别代表在浅层、中层、深层和特征拼接阶段所施加的分类损失。

2.1 ResNet 主干网络

在图像识别领域，基于卷积神经网络的识别网络结构层数越深，越能更好地学习到图像中的细节特征，识别结果得到不断提升，然而更深的网络具有更多的网络参数，并且网络训练的梯度消失问题更明显，训练误差加大。为解决此问题，He 等[25] 提出了一种残差网络 (residual network, ResNet) 结构，通过残差表示和快速链接解决了梯度消失问题。由于深层特征蕴含丰富的语义信息，因此本文选择 ResNet50 作为骨干网络。ResNet50 包含 1 层卷积层 (Conv1) 和 4 层残差结构 (Conv2~Conv5)，然而 ResNet50 浅层特征的提取能力较弱，不能很好地提取到车辆的细粒度特征。因此针对 ResNet50 选用后 4 层残差结构 (Conv2~Conv5) 作为骨干网络提取特征的 4 个阶段 (layer1~layer4)，网络结构如图 1 所示。

2.2 渐进式多粒度局部卷积模块

由于不同车型外观的局部差异细微，也就是说，大多数情况下不同的车型具有相同的全局特征，而只在细微的局部细节上有差异，因此车辆的局部细节特征往往比全局特征具有更强的车型特征辨别力，并且这些具有辨别力的局部区域面积差异较大，导致辨别性区域经常不在一个粒度层次上，因此，对于这种粒度差别较大的特征图像，常规网络注意力常关注到较大粒度的图像特征，容易忽略这种有辨别力的小粒度特征。例如，车灯相对车标的面积差异较大，那么，卷积神经网络更容易捕捉到粒度较大的“车灯”、“栅格”等区域特征，却难以关注到粒度较小的更具辨别力的“车标”、“车门把手”等区域特征。基于拼图生成器的渐进式多粒度网络 (progressive multi-granularity training of jigsaw patches, PMG)[26] 提出了一种用于细粒度视觉分类的新框架，利用拼图生成器生成不同粒度级别信息的车辆图像，在车型识别任务中取得较好的识别效果，然而由于这种策略需要将原图进行随机拼接，再将拼接后的图像送入网络进行训练，因此破坏了图像空间结构特征，使得特征学习复杂化，导致网络无法捕获真正有意义的车型识别区域。为解决此问题，提出一种渐进式多粒度局部卷积模块 (PLCB)，利用这种渐进式多粒度特征提取方式可以使得网络快速准确提取不同粒度大小的车辆局部特征，从而有效提高车型识别的准确性和鲁棒性。所提 PLCB 结构如图 2 所示。卷积操作提取特征，最后将所有提取完特征的局部图像在空间维度上进行拼接，因此有效保留了原有图像的空间结构特征。具体渐进式多粒度局部卷积模块 (PLCB) 的实现步骤如下:

首先，假设输入特征图 F_{PLCB}^{in} 的尺寸为 $B\times C\times H\times W$，其中 B 代表批次 (batch) 维度，C 代表通道 (channel) 维度，H 和 W 分别代表特征图所对应的高和宽。在不同阶段将网络在空间维度上进行切分，切分为不同大小的局部特征图，将切分后的特征图按顺序在批次 (batch) 维度进行拼接，得到 $F_{Split}^{i}\in 2^{2(4-i)}B\times\frac{H}{2^{2(4-i)}}\times\frac{W}{2^{2(4-i)}}\times C$，具体数学描述如式 (1) 所示：

$$F_{Split}^{i}=Concat\left[Split(F_{PLCB}^{in})\right],(i=1,2,3,4), \tag{1}$$

式中：$Split(\cdot)$ 代表切分操作，$Concat(\cdot)$ 代表拼接操作，i 代表训练的不同阶段，i 的取值为 1，2，3，4。

其次，将切分后的特征 F_{Split}^{i} 进行卷积操作特征提取，具体数据描述如式 (2) 所示:

$$F_{Conv}^{i}=Conv\left(F_{Split}^{i}\right),i=1,2,3,4, \tag{2}$$

式中，$Conv(\cdot)$ 代表卷积操作，用于主干网络特征提取的 4 个阶段进行卷积提取特征，其中所有切分后的局部卷积块共享同一权重参数。

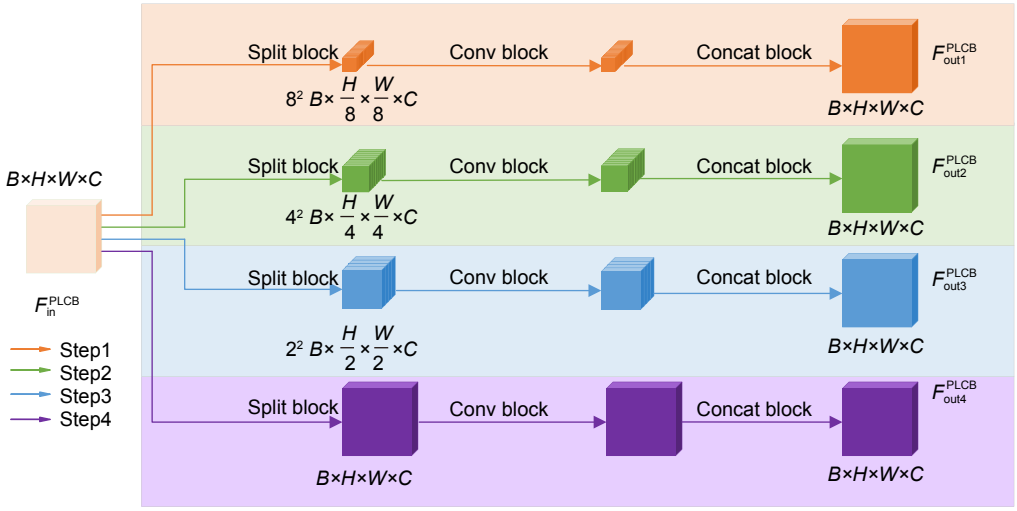


图 2　渐进式多粒度局部卷积模块 (PLCB)

Fig. 2　Progressive multi-granularity Local Convolution Block

最后，将卷积后的特征F^i_{Conv}在不同阶段将网络在批次 (batch) 维度上进行切分，将切分后的特征图按顺序在空间维度进行拼接，得到拼接后的特征$F^i_{\text{Concat}}\in B\times H\times W\times C$，具体数学描述如式 (3) 所示：

$$F^i_{\text{Concat}}=Concat\left[Split(F^i_{\text{Conv}})\right],(i=1,2,3,4),\quad(3)$$

式中：$Split(\cdot)$代表切分操作，$Concat(\cdot)$代表拼接操作，i代表训练的不同阶段，i的取值为 1，2，3，4。

由上述渐进式多粒度局部卷积模块 (PLCB) 的实现步骤可知，在 PLCB 模块中，当网络进行卷积操作时，不是直接在原图上进行卷积，而是对每个切分后的局部图像在批次 (batch) 维度拼接后进行局部卷积操作，不需要将图像进行打乱重组送入网络进行训练。因此，所提 PLCB 不仅能有效利用切分后的局部图像进行局部卷积操作，提取更多的多粒度车辆特征，而且可以更好保留车辆的空间结构特征。

2.3 随机通道丢弃模块

PLCB 将车辆特征图进行多粒度切分，并对切分后的每个局部图像进行不同粒度的局部卷积，能够快速有效定位具有判别力的局部特征区域，然而由于常规分类器往往为了提高分类精度仅关注最具鉴别力的特征，例如车灯是最具鉴别能力的车型识别特征，那么网络的注意力更多关注在如何利用车灯特征进行车型识别，但是在车型识别任务中不仅需要最具鉴别能力的车灯、栅格等车辆显著性特征，还需要充分利用车标、车门把手等其它非显著性特征提高车型识别的准确率。在以往的工作中，车型识别任务中大多使用注意力机制[16-18]的策略定位到车型所在区域，进而进行识别，然而基于注意力机制的的车型识别网络往往只会关注到一些较大粒度的显著性特征，忽略了其它微小粒度的细节特征。因此，受 Bilinear-CNN[19] 中利用归一化余弦距离测量通道相关性的启发，提出一种随机通道丢弃模块 (random channel drop block, RCDB)，所提 RCDB 模块利用双线性池化度量的方式生成通道相关矩阵表征各个通道之间的成对相关性，消除通道之间的共同适应关系，再引入 ADL[27] 中的 Dropout 操作进行随机信息丢弃，以鼓励网络更多关注除显著性区域外的其它具有辨别力的细粒度区域特征，然后利用随机通道丢弃操作抑制多粒度渐进式训练产生的局部最优过拟合现象。

由于所提 RCDB 模块采用随机方式屏蔽一组相关通道，使得车型识别具有更多的遮挡组合，提取车型特征具有更多的可能性，因此所提网络能关注到如车辆轮廓、车窗、车灯、车栅格等常规网络不易关注到的局部细粒度区域。所提网络利用 ResNet 网络中的残差模块对切分后的不同粒度车辆特征进行卷积操作，然后引入 RCDB 模块，通过随机选择局部图像中的部分特征进行置信度排序，通过将置信度排序靠前的置为 0，排序靠后的置为 1，从而抑制所有局部图像中的显著性区域，分散网络关注显著特征的注意力，诱导网络捕捉各种非显著性判别区域，具体操作如图 3 所示。实现步骤如下:

首先，假设输入特征图$F^{\text{in}}_{\text{RCDB}}$的尺寸为$C\times H\times W$，

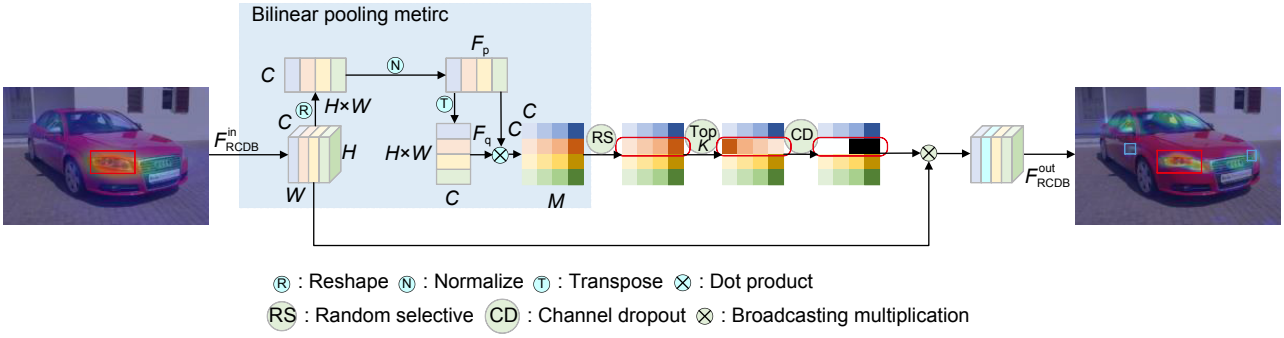


图 3 随机通道丢弃模块 (RCDB) 结构图

Fig. 3 Random channel drop block schematic diagram

经过 Reshape 操作将特征图的维度转换为$C\times HW$，在空间维度上进行归一化操作得到$F_{\rm p}$，其在空间维度HW上取值范围为 0~1，具体数据描述如式 (4) 所示：

$$F_{\rm p}=Normalization\left[Reshape(F_{\rm RCDB}^{\rm in})\right]\in C\times HW, \quad (4)$$

式中：$Reshape(\cdot)$代表变维操作，用于将二维矩阵伸展为一维矩阵；$Normalization(\cdot)$代表归一化操作。

然后，对归一化后的特征图进行转置操作得到$F_{\rm q}$，尺寸转换为$HW\times C$，得到具体数据描述如式 (5) 所示：

$$F_{\rm q}=(F_{\rm p})^{\rm T}\in HW\times C, \quad (5)$$

式中，$(\cdot)^{\rm T}$代表转置操作 (transpose)。

最后，将归一化后的特征$F_{\rm p}$与转置后的特征$F_{\rm q}$进行双线性池化 (bilinear pooling, BP) 操作，得到通道相关矩阵$\boldsymbol{M}\in R^{C\times C}$，用于描述每个特征通道之间的成对相似性，具体数学描述如式 (6) 所示：

$$\boldsymbol{M}=BP(F_{\rm p},F_{\rm q})\in C\times C, \quad (6)$$

式中，$BP(\cdot)$代表双线性池化操作。

由于每一个通道代表着识别车型的不同部位，因此通过建立通道相关矩阵 $\boldsymbol{M}$ 可以观察到不同通道之间的相关性。为了使网络提取到更丰富的车型特征，拥有更多的遮挡可能性，随机选择 $\boldsymbol{M}$ 中的一行，对选中的一行进行置信度排序，置信度排名前 k 个元素为显著性区域，排名 k 个元素之后的为非显著性区域。如果设置显著性区域为 0，其它非显著性区域为 1，利用广播乘法将下降掩码应用到输入特征映射中，最终置信度排名前 k 个相邻区域的特征被一起抑制，即网络丢弃最显著区域，迫使网络去关注其它非显著性区域。k 的大小由通道丢弃比例β决定，具体数学描述如式 (7) 所示:

$$C_k=\begin{cases}0, & \text{if} \quad \boldsymbol{M}_i>I_k\\ 1, & \text{elsewise}\end{cases}, \quad (7)$$

式中：C_k代表丢弃掩码 (crop mask)，i 代表通道相关矩阵中的任意一行，$\boldsymbol{M}_i$代表第 i 行的通道相关矩阵，I_k代表排名第 k 个元素所含信息量，k 的大小为$c\times\beta$，其中 c 代表通道数，β代表通道丢弃比例。

2.4 渐进式多粒度训练模块

渐进式多粒度训练模块 (Progressive multi-granularity training block, PMTB) 将主干网络分为不同的阶段，对不同阶段提取到的特征施加分类损失并进行参数训练更新，然后将每个阶段训练的参数输入到下一阶段。利用所提的 PMTB 以及渐进式多粒度局部卷积模块 (PLCB) 有效融合多粒度车辆特征，使其在不同阶段的训练过程中联合协作，从而能显著提高所提网络对车辆不同粒度特征的捕获能力。多粒度渐进式训练 4 个阶段示意图如图 4 所示。特征提取网络可以被分为 L 个阶段，每个阶段的输出为F^l，通过不同阶段的网络组合，先后接入卷积模块 (Conv block)、多层感知机 (multilayer perceptron, MLP)，其中 MLP 由 2 个批量归一化 (batch normalization, BN)、1 个 ReLU 激活函数和 2 个全连接层 (fully connected layers, FC) 层构成，最终得到各层的分类预测概率y^l，具体数学描述如式 (8)、式 (9) 所示：

$$V^l=H_{\rm Conv}^l(F^l), \quad (8)$$

$$y^l=H_{\rm Class}^l\left(V^l\right). \quad (9)$$

最后，将 S 个阶段的网络输出建立联合特征表示层$V^{\rm Concat}$，并接入 MLP，得到拼接浅层、中层和深层特征的分类预测概率$y^{\rm Concat}$，具体数学描述如式 (10) 所示：

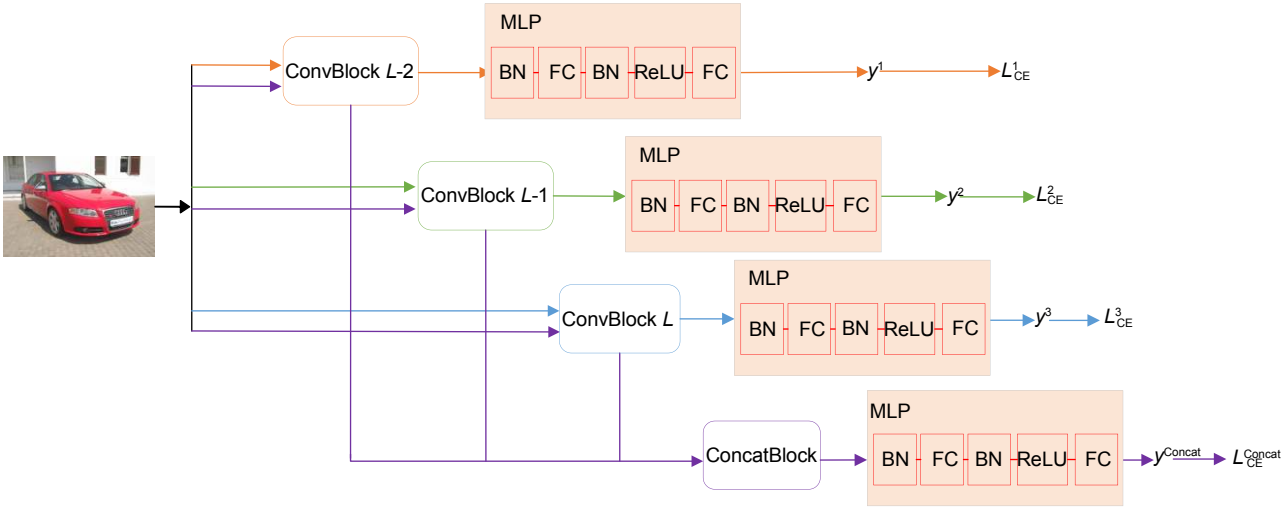


图 4 渐进式多粒度训练模块 (PMTB) 示意图

Fig. 4 Progressive multi-granularity training block schematic diagram

$$y^{\text{Concat}} = H_{\text{class}}^{\text{Concat}}\left\{\text{Concat}[V^{(L-S+1)},\ldots,V^{L}]\right\}. \tag{10}$$

在模型的训练过程，每一次训练迭代过程包括 S+1 个步骤，每一步只训练一次的输出，每次迭代将整个网络的参数进行更新，使得网络可以学习车辆的多粒度特征信息。多粒度网络学习应用于 PMTB 的训练步骤如表 1 所示。

表 1 渐进式多粒度训练步骤

Table 1 Progressive multi-granularity training steps

渐进式多粒度训练过程
输入：训练数据集D，训练数据的批次为x，标签样本为y，P代表多粒度渐进式网络学习，L_{CE}代表交叉熵损失(cross entropy loss, CE)
For $epoch \in [0, epochs]$ do
For $b \in [0, batchs]$ do
$x, y \Leftarrow batch\ b$ of D
For $l \in [L-S+1, L]$ do
$y^l \Leftarrow H_{\text{class}}^l[H_{\text{Conv}}^l(F^l(P(x,n))))]$
$L_l \Leftarrow L_{CE}(y^l, y)$
Backpropagation L_l
End for
$y^{\text{Concat}} = H_{\text{class}}^{\text{Concat}}\left\{\text{Concat}[V^{(L-S+1)},\ldots,V^{L}]\right\}$
$L_{\text{Concat}} \Leftarrow L_{CE}(y^{\text{Concat}}, y)$
Backpropagation L_{Concat}
End for
End for

2.5 损失函数

损失函数是决定卷积神经网络模型性能的关键因素之一。交叉熵 (CE) 损失函数是车型识别模型常用的损失函数，采用交叉熵损失函数进行损失计算能够有效反映车型识别的类间特征差异，交叉熵损失函数数学描述如式 (11)、(12) 所示：

$$L_{CE}^{i}(y_i^l, y_i) = -\sum_{i=1}^{m} y_i^l \times \log(y_i)(i = 1,2,3), \tag{11}$$

$$L_{CE}^{\text{Concat}}(y_i^l, y_i) = -\sum_{i=1}^{m} y_{\text{Concat}}^l \times \log(y_{\text{Concat}}), \tag{12}$$

式中，L_{CE}^{i}代表第 i 个阶段的输出；L_{CE}^{Concat}代表第 4 个阶段的输出；y_i代表车型的真实标签，y_i^l代表车型的预测标签，m 代表一个批次送入模型训练的图像数量。

由于每个阶段的预测中所有参数在上一个阶段已经得到更新，因此每个阶段的损失仅为针对当前阶段的预测。整个网络的总损失 L_{Loss} 为 4 个阶段相加的损失之和。网络总交叉熵损失数学描述如式 (13) 所示：

$$L_{\text{Loss}} = L_{CE}^{1} + L_{CE}^{2} + L_{CE}^{3} + L_{CE}^{\text{Concat}}, \tag{13}$$

式中，L_{Loss} 代表网络总的交叉熵损失。

3 实验结果与分析

实验平台采用 Inter(R) Silver 4210R 处理器，128 G 内存，GPU 为 Nvidia RTX3090，显存为 24 GB；深度学习框架选用 Pytorch1.7.0 与 CUDA11.4 的 GPU 运行版本以及 cuDNN11.0 深度学习 GPU 加速库。实验数据集采用 Stanford-cars 数据集[28]、Compcars 网络数据集[29] 和真实场景下的车型数据集 VMRURS[30]，

其中 Stanford-cars 数据集由斯坦福-人工智能实验室发布，共包含 196 种车型的图像数据，16185 张图像，其中训练集和测试集分别为 8144 张和 8041 张，每张图片的标签包含制造商、车辆型号和生产年份 3 个信息。Compcars 数据集包含 2 类，即卡口监控数据集和网络数据集，卡口数据集中的图像车辆姿态固定，因此较低了分类难度，为有效地体现本文方法的优势，选用 Compcars 网络数据集，数据集中的车辆具有各种不同的姿态，共包含 163 种大类车型，1993 种小类车型，共 30955 张图像，其中训练集和测试集分别为 16016 张和 14939 张。真实场景下的车型数据集 VMRURS 不同于上述 2 种数据集来自于网络爬取，它采集于安装在高速公路上的不同视角下、不同帧率下的摄像机所捕捉到的高分辨率视频，有助于进一步验证基于真实场景下的车型识别效果，此外，为了维护数据集中的个人隐私和保护车主身份，将车牌字符和车主个人面部都进行了手动模糊。该数据集共 3847 张图像，包含 48 种车型，训练集和测试集分别为 3096 张和 751 张图像，数据集来自于不同光照强度、不同视角下的高速公路上捕捉到的图像。

车型识别评价指标采用分类准确率 A (accuracy) 进行评价，评价公式数学描述如式 (14) 所示：

$$A=\frac{\sum_{i=1}^{m}\widetilde{y}_i=y_i}{m}, \tag{14}$$

其中：i 表示车辆样本序号，m 为车辆总数，$\widetilde{y}_i$表示车型识别模型预测输出，y_i为车型真实标签。

3.1 网络参数设置

对于所提的渐进式多粒度 ResNet 车型识别网络，在训练阶段只引入类别标签，输入图像尺寸调整为 550×550，并随机裁剪尺寸为448×448，采用水平翻转进行数据增强，在验证阶段输入图像尺寸调整为 550×550，并中心裁剪尺寸为448×448，进行预测时只是用y^{Concat}进行预测，不需要使用前三个阶段的预测结果。所提出的 PLCB 模块和 RCDB 模块只在训练阶段起作用，在验证阶段不涉及额外的参数和计算成本。整个训练过程中批次大小设置为 24，共学习 200 轮。初始学习率设置为5×10^{-4}，在第 100 轮和第 150 轮分别衰减为5×10^{-5}和5×10^{-6}，动量设置为 0.9，学习率下降策略使用余弦退火策略进行梯度下降，优化器选用随机梯度下降 (stochastic gradient descent, SGD)，通道丢弃比例β设置为 0.25。图 5 给出了在 Stanford-cars 数据集、Compcars 网络数据集和真实场景下的车型数据集 VMRURS 上，top1 准确率随β(0-0.4) 取值的变化曲线图，也验证了β设置为 0.25 的合理性。

从图 5 可以看出，当超参数β选取为 0.25 时，网络的性能最优，此时网络引入 RCDB 模块，通过随机通道丢弃方法不仅可以使网络关注到显著性区域，还可以关注到非显著性区域。当$0\leqslant\beta\leqslant0.2$时，由于丢弃比例过低，网络无法掩盖显著性区域，减弱了网络关注非显著性区域的能力，因此网络的性能出现了下滑。当$0.3\leqslant\beta\leqslant0.4$时，由于丢弃比例过高，会掩盖掉过多的显著性特征，不利于最终的识别效果，所以网络的性能同样出现了下滑，所以超参数β的较优取值区间应该为[0.2，0.3]，实验中β统一选择为 0.25。

3.2 网络训练和测试

对于所提出的多粒度渐进式 ResNet 车型识别网络，在 Stanford-cars 数据集中进行训练和测试时的准确率和损失变化情况如图 6 所示。从图中可以看出，随着训练轮次的增加，识别准确率不断提高，损失值

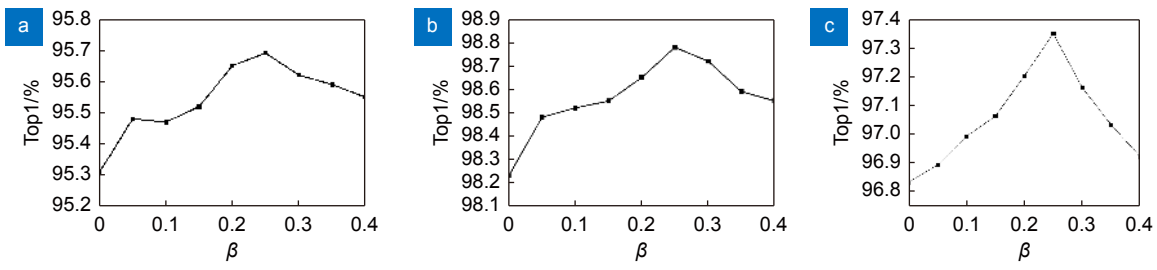


图 5 Top1/% 变化曲线图。(a) Stanford-cars 上 β 值对 RCDB 的影响；(b) Compcars 上 β 值对 RCDB 的影响；(c) VMRURS 上 β 值对 RCDB 的影响

Fig. 5 Top1/% curve of change. (a) Effect of β values on RCDB on Stanford-cars; (b) Effect of β values on RCDB on Compcars; (c) Effect of β values on RCDB on VMRURS

不断降低。当训练损失和验证损失趋于稳定时，训练准确率和验证准确率逐渐收敛，训练准确率 train_Acc 最终达到了 99.8%，验证准确率 test_Acc 最终达到了 95.7%。

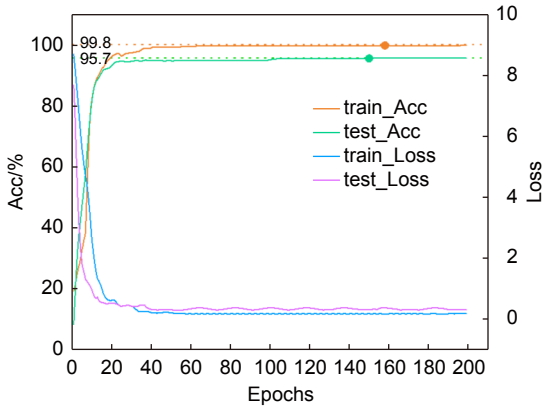


图 6 网络训练和验证过程

Fig. 6 Network training and testing process

3.3 消融实验对比分析

3.3.1 PLCB 模块切分对比实验

为验证渐进式多粒度 ResNet 车型识别网络各模块的作用，在 Stanford-cars 数据集上开展了消融实验，表 2 给出了 PLCB 在不同阶段网格切分大小对网络准确率影响的比较。通过实验比较图像有无网格切分对最终分类效果的影响，并利用网络切分参数 n 的取值分析网络性能的差别，对比结果如表 2 所示。

表 2 各个阶段 PLCB 切分大小的比较

Table 2 Comparison of PLCB split size at each stage

Stage1	Stage2	Stage3	Stage4	Accuracy/%
1	1	1	1	93.5
2	2	2	2	93.9
4	**4**	**4**	**4**	**94.2**
8	8	8	8	94.0
16	8	4	2	93.6
8	**4**	**2**	**1**	**94.5**
4	2	1	1	93.9

表 2 中，第 1 行 n=1 代表没有进行网格切分操作时的识别准确率，第 2、3、4 行分别代表将图像切分成 2×2、4×4、8×8 大小的网格。由表 2 可以看出，引入网格切分操作的识别效果优于无网格切分的效果，表明渐进式局部卷积模块对车型识别的有效性；当 n=2 时，网格切分过粗，会导致提取局部特征效果不明显；当 n=8 时，网格切分过细，会导致产生过多的空白网格，引入过多无效特征，造成识别效果有所下降；当 n=4 时，通过网格化的方法将每个阶段的感受野限制在原始大小的 1/4，此时识别效果最佳。第 5、6、7 行结合多粒度思想，以指数递减的方式对参数 n 进行设置。也就是说，在训练网络的不同阶段 n 是不同的。大量实验结果表明，当$n=\{16,8,4,2\}$时，由于网格切分过于精细，网格对过于精细的区域识别效果有限，导致识别效果有所下降；当$n=\{4,2,1,1\}$，由于网络切分过于粗糙，网络在浅层还无法实现较为精细的特征，无法有效提取到细粒度车型特征，造成识别效果不佳；当$n=\{8,4,2,1\}$时，网络能够基于已学习到的细粒度模式去学习粗粒度模型，因此具有最佳识别效果，也验证了多粒度策略对车型识别的有效性。

3.3.2 RCDB 插入位置选择

为了进一步对比所提的 RCDB 模块加入网络的不同层数对车型识别任务的有效性，量化分析了加入 RCDB 模块的不同位置对车型识别效果的影响，对比实验结果如表 3 所示。对比该模块加入 ResNet50 中 4 个残差网络不同层的输出后的模型识别效果。由表 3 可以看出，相比原始的 ResNet50，提出的 RCDB 模块加入不同层之后的识别效果都有不同程度的提升，其中在 Layer2 和 Layer3 层的提升效果最明显，分别提升了 1.4% 和 1.5%，这是因为中间层兼顾了深层语义信息和浅层细节，能提取到更具辨别力的多粒度车型识别特征。相比于单层模块，在每层后都加入 RCDB 模块具有更好的识别效果。当 ResNet50 在 Layer1-Layer4 中均加入该模块后，其识别效果达到最佳。

表 3 RCDB 模块加入不同层后识别效果消融实验

Table 3 Ablation experiment of recognition effect after adding different layers to the RCDB module

ResNet50	Layer1	Layer2	Layer3	Layer4	Accuracy/%
✓	✗	✗	✗	✗	91.5
✓	✓	✗	✗	✗	92.3
✓	✗	✓	✗	✗	92.9
✓	✗	✗	✓	✗	93.0
✓	✗	✗	✗	✓	92.6
✓	✓	✓	✓	✓	**93.2**

3.3.3 模型有效性实验

为了进一步验证所提网络及其各个模块对车型识

别任务的有效性，在 Stanford-cars 数据集上进行消融实验分析，实验结果如表 4 所示。从表中可以看出，Baseline 由于其强大的特征提取能力，在车型识别任务中已经取得了比较好的结果，准确率已经达到 91.5%；在 Baseline 的基础上添加渐进式多粒度局部卷积模块 (PLCB) 后，使得网络准确率提高了 3.3%；添加随机通道丢弃模块 (RCDB) 模块，网络准确率提高了 1.7%；最后将 PLCB 和 RCDB 加入 Baseline 中，所提网络与 Baseline 相比提高了 4.2%，达到了 95.7%。因此，消融实验对比结果验证了所提网络的有效性。

表 4　不同模块依次加入网络中的实验效果

Table 4　Different modules are added to the network

Baseline	PLCB	RCDB	Accuracy/%
✓	✗	✗	91.5
✓	✓	✗	94.8
✓	✗	✓	93.2
✓	✓	✓	**95.7**

3.4 定性实验对比分析

3.4.1 PLCB 各阶段可视化对比

为了进一步验证渐进式多粒度 ResNet 车型识别网络的有效性，利用渐进式多粒度局部卷积模块 (PLCB) 和渐进式多粒度训练模块 (PMTB) 得到各个阶段的特征图，进行可视化实验对比分析。具体地，在 Stanford-cars 数据集、Compcars 网络数据集和真实场景下的车型数据集 VMRURS 上选择不同视角下的车辆进行可视化。网络在不同阶段提取到车辆不同的多粒度特征可视化结果如图 7 所示。图 7 中 Step1 表示 PLCB 提取的浅层特征，参数 n 设置为 8；Step2 表示 PLCB 提取的中层特征，参数 n 设置为 4；Step3 表示 PLCB 提取的深层特征，参数 n 设置为 2；Step4 表示 PLCB 将浅层、中层和深层特征拼接后的特征，参数 n 设置为 1。从图 7 可知，在 Step1 阶段，由于模型感受野受限，PLCB 对车辆的复杂特征描述能力不足，因此只能关注到车辆纹理信息或者背景噪声信息；在 Step2 阶段，PLCB 既能提取到一部分高层语义信息，也能提取一部分低层纹理信息，因此网络不能准确定位到显著性区域，会关注到部分车辆特征；在 Step3 阶段，PLCB 在较深层有较强的特征提取能力，网络可以定位到显著性判别区域，如车灯等信息。在 Step4 阶段，PLCB 拼接前 3 个阶段提取到的特征，由于深层特征具有很强的语义信息，图像表征能力较强，在特征提取部分占主导性地位，因此在 Step4 阶段提取到的特征在可视化效果上与 Step3 阶段有很强的一致性，通过可视化效果可表明通过渐进式多粒度训练模块 (PMTB) 对所提网络进行训练，网络可以准确定位到显著性判别区域。

3.4.2 加入各模块后可视化对比

为了进一步验证渐进式多粒度 ResNet 车型识别网络各个模块的有效性，在 Stanford-cars 数据集、Compcars 网络数据集和真实场景下的车型数据集 VMRURS 上选择不同视角下的车辆依次加入各个模块并进行可视化分析，实验结果如图 8 所示。第二行使用 Baseline 提取特征后，所提网络的高激活区域分布范围主要在显著性区域，但由于网络特征提取能力不足，经常出现定位不准以及背景噪声干扰，严重影响车型识别效果。如图 8(b) 中识别到后尾灯，但是也识别到其他干扰信息，图 8(c) 中识别到车前灯处，但是也识别到其他干扰车辆，图 8(g) 中识别到车灯处，

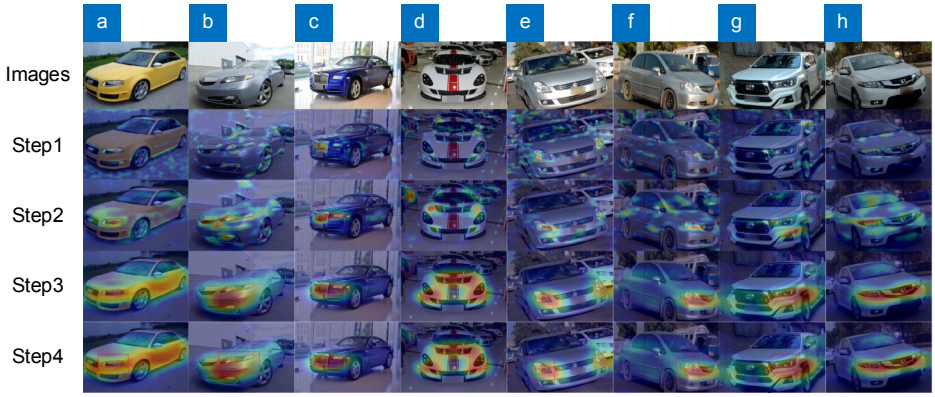


图 7　各个阶段的车型识别可视化对比

Fig. 7　Visual comparison of vehicle recognition in each stage

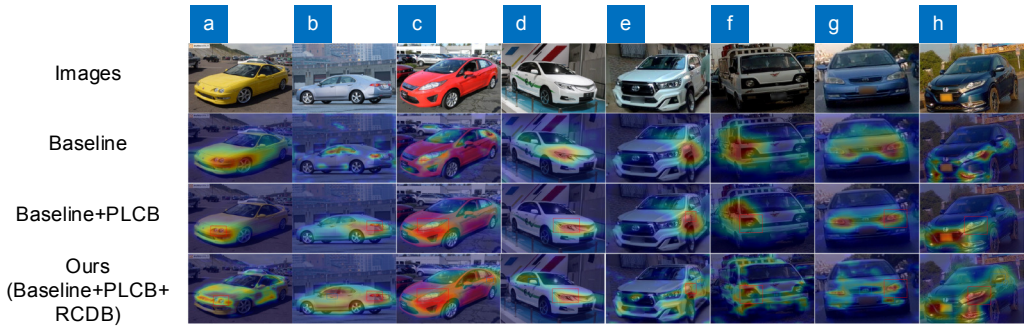


图 8　加入各模块后可视化对比

Fig. 8　Visual comparison of after adding each module

但是也识别到车身机盖对识别没有帮助的区域。第三行加入 PLCB 模块后，所提网络通过多粒度局部卷积操作通过多阶段引导能够快速准确定位到显著性区域。如图 8(b) 中可以准确定位到车尾灯显著性特征进行识别，图 8(c) 中可以准确定位到车前灯处进行识别，图 8(g) 中可以准确识别到车灯处。第 4 行为本文所提网络，即当同时加入 PLCB 模块和 RCDB 模块后，所提网络不仅可以观察到车辆的显著性特征区域，还可以观察到其它有辨别力的非显著性区域，融合车辆的显著性特征和有辨别力的非显著性特征，如图 8(b) 中不仅可以识别到车尾灯等显著性特征，还可以识别到车门把手等非显著性特征，图 8(c) 中不仅可以识别到车灯信息，还可以识别到车后门把手，图 8(g) 中不仅可以识别到车灯特征，还可以识别到车标等微小特征，进而能够有效提高车型识别准确率。

3.4.3 不同网络可视化对比

为了进一步证明渐进式多粒度 ResNet 车型识别网络在不同干扰条件下车型识别的鲁棒性，选取易分类错误的车辆图像进行可视化对比分析，对比结果如图 9 所示。实验样本包含：1) 拍摄角度的影响因素，如第 1 行所示，车辆图像是自上而下拍摄的，并且在训练集中缺少相同拍摄角度的样本；2) 车辆本身发生形变，如第 2 行所示，车辆四个车门敞开，使得车身部分可以提取到的有效特征发生了变化，对识别造成干扰；3) 车辆本身在原图中所占比例过小，背景所占比例过大，如第 3 行所示；4) 环境光照等恶劣情况对识别造成干扰，如第 4 行所示，光照强度过低，无法提取到有效特征；5) 原图中有多个车辆，对识别造成干扰，如第 5 行所示。对比网络为近两年现有的基于弱监督学习策略的细粒度车型识别网络，包括 3 种基于多尺度特征的策略，分别为细粒度特征增强抑制网络 (feature boosting, suppression, and diversification, FBSD)，细粒度跨层多尺度网络 (cross-X learning for fine-grained visual categorization, Cross-X)，细粒度跨层引导网络 (cross-layer navigation convolutional neural network, CN-CNN)；2 种基于注意力学习的策略，分别为细粒度注意力增强网络 (weakly supervised data augmentation network, WS-DAN)，细粒度反事实注意力网络 (counterfactual attention learning, CAL)；1 种基于目标定位的策略，细粒度结构建模网络 (look-into-object: self-supervised structure modeling for object recognition, LIO)；2 种特征重构的策略，分别为细粒度破坏重构网络 (destruction and construction learning for fine-grained image recognition, DCL)，基于拼图生成器的多粒度渐进式网络 (progressive multi-granularity training of jigsaw patches, PMG)。由实验对比结果可以看出，FBSD[31]、LIO[32]、DCL[33]、Cross-X[34]、CAL[17]、WS-DAN[18]、PMG[26]、CN-CNN[35] 等对比网络在不同干扰下进行车型识别时效果都较差，对比网络常出现误识别、无法聚焦到有效区域等问题，而所提网络通过引入渐进式多粒度局部卷积模块 (PLCB) 和随机通道丢弃模块 (RCDB)，不仅可以快速准确地定位到显著性区域，还能通过抑制显著性区域，使网络更多关注到其它非显著性区域，因此在车型识别效果上明显优于对比网络。

3.5 定量实验对比分析

为定量分析提出的渐进式多粒度 ResNet 车型识别网络有效性，与现有的一些先进的基于弱监督学习策略的车型识别方法在 Stanford-cars 数据集[28]、Compcars 网络数据集[29] 和真实场景下的车型数据集 VMRURS[30] 上进行比较。具体的，对比网络包括 FBSD[31]、LIO[32]、DCL[33]、Cross-X[34]、CAL[17]、WS-

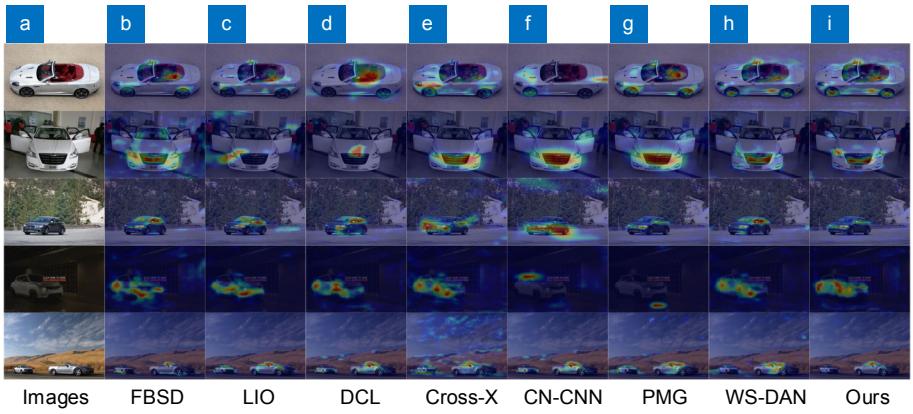


图 9 不同网络车型识别可视化对比

Fig. 9 Visual comparison of different network vehicle recognition

DAN[18]、PMG[26] 和 CN-CNN[35]。上述提到的所有网络都是在基线网络 (ResNet50) 的基础上进行改进，其中 FBSD[31] 引入了两个轻量化模块，一方面增强特征图中最显著的部分，另一方面学习车辆图像语义互补信息，最后捕捉到多尺度特征信息，识别效果在三个数据集上分别提升到 94.4%、96.8% 和 92.3%；LIO[32] 通过自监督的方式，对车型的整体结构进行建模，有效利用了车辆的内部结构特征，识别效果在三个数据上分别提升到 94.5%、96.8% 和 94.2%；Cross-X[34] 利用不同图像以及不同网络层之间的关系来进行稳健的多尺度特征学习，识别效果在三个数据集上分别提升到 94.6%、97.0% 和 94.6%；DCL[33]、PMG[26] 采用不同的策略将原图进行打乱，然后结合多粒度思想提取车辆的多粒度局部特征，在三个数据集上识别效果分别达到了 94.5%、96.7%、94.7% 和 95.1%、97.8%、95.7%；CAL[17] 和 WS-DAN[18] 通过引入注意力机制，将显著性区域进行放大，在三个数据集上识别效果分别达到了 95.5%、98.0%、96.4% 和 94.5%、97.1%、95.6%；CN-CNN[35] 利用金字塔结构和长短期记忆网络，设计出一种自顶向下和自底向上的双向特征传递路径，实现不同层之间的特征交互，在三个数据集上识别效果分别达到了 94.9%、97.6% 和 94.9%。本文所提出的渐进式多粒度 ResNet 网络基于多粒度渐进式训练的策略，通过利用渐进式多粒度局部卷积模块 (PLCB) 结合随机通道丢弃模块 (RCDB) 提取到车辆的多样性、有辨别力的多粒度特征，在三个数据集上识别效果分别达到了 95.7%、98.8% 和 97.4%，优于其它车型识别对比网络。具体对比结果如表 5 所示。

表中 Params 表示参数量，能够衡量模型的空间复杂度，FLOPs (floating-point operations) 表示浮点运行次数，能够衡量模型的时间复杂度，表中 Speed 表示模型运行速度。由于表中所有模型都是在基线[25] 网络的基础上进行改进，因此导致在 Params 和

表 5 不同网络车型识别准确率比较

Table 5 Comparison of recognition accuracy of different network models

Methods	Backbone	Stanford-cars/%	Compcars/%	VMRURS/%	Speed/(f/s)	Params/M	FLOPs/G
基线[25]	ResNet50	91.5	94.1	87.1	**4.15**	**23.50**	**33.05**
FBSD[31]	ResNet50	94.4	96.8	92.3	1.73	46.82	53.11
LIO[32]	ResNet50	94.5	96.8	94.2	3.60	24.57	33.06
DCL[33]	ResNet50	94.5	96.7	94.7	3.46	24.91	33.06
Cross-X[34]	ResNet50	94.6	97.0	94.6	3.88	25.56	38.86
CAL[17]	ResNet50	95.5	98.0	96.4	3.72	33.73	33.08
WS-DAN[18]	ResNet50	94.5	97.1	95.6	4.02	33.24	33.08
PMG[26]	ResNet50	95.1	97.8	95.7	2.94	45.12	69.82
CN-CNN[35]	ResNet50	94.9	97.6	94.9	1.92	42.31	47.65
Ours	ResNet50	**95.7**	**98.8**	**97.4**	2.97	40.64	69.61

FLOPs 的计算上都有不同程度的增加。与其他 8 种基于弱监督学习策略的车型识别算法相比，所提模型由于引入了 PLCB 和 RCDB 模块，并使用 PMTB 的训练模块进行训练，导致与其他模型相比，训练的参数量有所提升，达到了 40.64 M，模型的计算量也有所提升，达到了 69.61 G。但是与同样使用多粒度思想的 PMG[26] 相比，模型在识别准确率提高的前提下，Params 和 FLOPs 都有不同程度的减少。在速度方面，本文的网络速度达到了 2.97 f/s 的速度，在基线网络的基础上，牺牲了部分推理速度，取得了更优异的识别准确率。

4 结束语

提出了一种渐进式多粒度 ResNet 车型识别网络，首先，以 ResNet 网络作为骨干网络，提出渐进式多粒度局部卷积模块 (PLCB)，对不同粒度级别的车辆图像进行局部卷积操作，使网络重构时能够关注到不同粒度级别的车辆局部特征；其次，对多粒度局部特征图利用随机通道丢弃模块 (RCDB) 进行随机通道丢弃，抑制网络对车辆显著性区域特征的注意力，提高非显著性车辆细粒度特征的关注度；最后，利用渐进式多粒度训练模块 (PMTB) 对所提网络进行训练，引导网络提取更具辨别性和多样性的特征。在 Stanford-cars 数据集[28]、Compcars 网络公开车型数据集[29] 和真实场景下的车型数据集 VMRURS[30] 上，相继开展了消融实验对比分析、定性实验对比分析、定量实验对比分析，所提模型在三个数据集上的识别准确率分别达到了 95.7%、98.8% 和 97.4%，充分证明了所提方法具有较高准确率和鲁棒性。在未来工作中，将进一步利用多粒度渐进式训练的策略，从粗粒度到细粒度提取多粒度特征，并考虑结合特征对齐思路使得模型获得更高的精度。

参考文献

[1] Bay H, Tuytelaars T, Van Gool L. SURF: speeded up robust features[C]//*Proceedings of the 9th European Conference on Computer Vision*, 2006: 404–417. https://doi.org/10.1007/11744023_32.

[2] Csurka G, Dance C R, Fan L X, et al. Visual categorization with bags of keypoints[C]//*Workshop on Statistical Learning in Computer Vision*, Prague, 2004.

[3] De Sousa Matos F M, De Souza R M C R. An image vehicle classification method based on edge and PCA applied to blocks[C]//*International Conference on Systems, Man, and Cybernetics*, 2012: 1688–1693. https://doi.org/10.1109/ICSMC.2012.6377980.

[4] Behley J, Steinhage V, Cremers A B. Laser-based segment classification using a mixture of bag-of-words[C]//*2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2013: 4195–4200. https://doi.org/10.1109/IROS.2013.6696957.

[5] Liao L, Hu R M, Xiao J, et al. Exploiting effects of parts in fine-grained categorization of vehicles[C]//*Proceedings of the 2015 IEEE International Conference on Image Processing*, 2015: 745–749. https://doi.org/10.1109/ICIP.2015.7350898.

[6] Hsieh J W, Chen L C, Chen D Y. Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition[J]. *IEEE Trans Intell Transp Syst*, 2014, **15**(1): 6−20.

[7] Feng J Z, Ma X C. Fine-grained entity type classification based on transfer learning[J]. *Acta Autom Sin*, 2020, **46**(8): 1759−1766.
冯建周, 马祥聪. 基于迁移学习的细粒度实体分类方法的研究[J]. 自动化学报, 2020, **46**(8): 1759−1766.

[8] Luo J H, Wu J X. A survey on fine-grained image categorization using deep convolutional features[J]. *Acta Autom Sin*, 2017, **43**(8): 1306−1318.
罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述[J]. 自动化学报, 2017, **43**(8): 1306−1318.

[9] Wang R G, Yao X C, Yang J, et al. Deep transfer learning for fine-grained categorization on micro datasets[J]. *Opto-Electron Eng*, 2019, **46**(6): 180416.
汪荣贵, 姚旭晨, 杨娟, 等. 基于深度迁移学习的微型细粒度图像分类[J]. 光电工程, 2019, **46**(6): 180416.

[10] Wei X S, Song Y Z, Aodha O M, et al. Fine-grained image analysis with deep learning: a survey[J]. *IEEE Trans Pattern Anal Mach Intell*, 2022, **44**(12): 8927−8948.

[11] Yang Z, Luo T G, Wang D, et al. Learning to navigate for fine-grained classification[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 438–454. https://doi.org/10.1007/978-3-030-01264-9_26.

[12] Fang J, Zhou Y, Yu Y, et al. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture[J]. *IEEE Trans Intell Transp Systems*, 2017, **18**(7): 1782−1792.

[13] Zhang X P, Xiong H K, Zhou W G, et al. Fused one-vs-all features with semantic alignments for fine-grained visual categorization[J]. *IEEE Trans Image Process*, 2016, **25**(2): 878−892.

[14] Xu H P, Qi G L, Li J J, et al. Fine-grained image classification by visual-semantic embedding[C]//*Proceedings of the 27th International Joint Conference on Artificial Intelligence*, 2018: 1043–1049. https://doi.org/10.5555/3304415.3304563.

[15] Zhang H, Xu T, Elhoseiny M, et al. SPDA-CNN: Unifying semantic part detection and abstraction for fine-grained recognition[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1143–1152. https://doi.org/10.1109/CVPR.2016.129.

[16] Ding Y F, Ma Z Y, Wen S G, et al. AP-CNN: weakly supervised attention pyramid convolutional neural network for fine-grained visual classification[J]. *IEEE Trans Image Process*, 2021, **30**: 2826−2836.

[17] Rao Y M, Chen G Y, Lu J W, et al. Counterfactual attention learning for fine-grained visual categorization and re-identification[C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 1005–1014. https://doi.org/10.1109/ICCV48922.2021.00106.

[18] Hu T, Qi H G, Huang Q M, et al. See better before looking closer:

weakly supervised data augmentation network for fine-grained visual classification[Z]. arXiv: 1901.09891, 2019. https://doi.org/10.48550/arXiv.1901.09891.

[19] Lin T Y, RoyChowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition[C]//*Proceedings of the 2015 IEEE International Conference on Computer Vision*, 2015: 1449–1457. https://doi.org/10.1109/ICCV.2015.170.

[20] Yu C J, Zhao X Y, Zheng Q, et al. Hierarchical bilinear pooling for fine-grained visual recognition[C]//*Proceedings of the 15th European Conference on Computer Vision*, 2018: 595–610. https://doi.org/10.1007/978-3-030-01270-0_35.

[21] Gao Y, Beijbom O, Zhang N, et al. Compact bilinear pooling[C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 317–326. https://doi.org/10.1109/CVPR.2016.41.

[22] Kong S, Fowlkes C. Low-rank bilinear pooling for fine-grained classification[C]//*Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 7025–7034. https://doi.org/10.1109/CVPR.2017.743.

[23] Sun M, Yuan Y C, Zhou F, et al. Multi-attention multi-class constraint for fine-grained image recognition[C]//*15th European Conference on Computer Vision*, 2018: 834–850. https://doi.org/10.1007/978-3-030-01270-0_49.

[24] Zheng X W, Ji R R, Sun X S, et al. Towards optimal fine grained retrieval via decorrelated centralized loss with normalize-scale layer[C]//*Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, 2019: 1140. https://doi.org/10.1609/aaai.v33i01.33019291.

[25] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 770–778. https://doi.org/10.1109/CVPR.2016.90.

[26] Du R Y, Cheng D L, Bhunia A K, et al. Fine-grained visual classification via progressive multi-granularity training of jigsaw patches[C]//*16th European Conference on Computer Vision*, 2020: 153–168. https://doi.org/10.1007/978-3-030-58565-5_10.

[27] Choe J, Shim H. Attention-based dropout layer for weakly supervised object localization[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 2214–2223. https://doi.org/10.1109/CVPR.2019.00232.

[28] Krause J, Stark J, Deng L, et al. 3D object representations for fine-grained categorization[C]//*2013 IEEE International Conference on Computer Vision Workshops*, 2013: 554–561. https://doi.org/10.1109/ICCVW.2013.77.

[29] Yang L J, Luo P, Loy C C, et al. A large-scale car dataset for fine-grained categorization and verification[C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015: 3973–3981. https://doi.org/10.1109/CVPR.2015.7299023.

[30] Ali M, Tahir M A, Durrani M N. Vehicle images dataset for make and model recognition[J]. *Data Brief*, 2022, **42**: 108107.

[31] Song J W, Yang R Y. Feature boosting, suppression, and diversification for fine-grained visual classification[C]//*International Joint Conference on Neural Networks*, 2021: 1–8. https://doi.org/10.1109/IJCNN52387.2021.9534004.

[32] Zhou M H, Bai Y L, Zhang W, et al. Look-into-object: self-supervised structure modeling for object recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11771–11780. https://doi.org/10.1109/CVPR42600.2020.01179.

[33] Chen Y, Bai Y L, Zhang W, et al. Destruction and construction learning for fine-grained image recognition[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 5152–5161. https://doi.org/10.1109/CVPR.2019.00530.

[34] Luo W, Yang X T, Mo X J, et al. Cross-x learning for fine-grained visual categorization[C]//*IEEE/CVF International Conference on Computer Vision*, 2019: 8241–8250. https://doi.org/10.1109/ICCV.2019.00833.

[35] Guo C Y, Xie J Y, Liang K M, et al. Cross-layer navigation convolutional neural network for fine-grained visual classification[C]//*ACM Multimedia Asia*, 2021: 49. https://doi.org/10.1145/3469877.3490579.

作者简介

徐胜军 (1976-)，男，陕西西安人，工学博士，副教授，硕士生导师，主要从事图像处理、模式识别领域的研究。

E-mail：duplin@sina.com

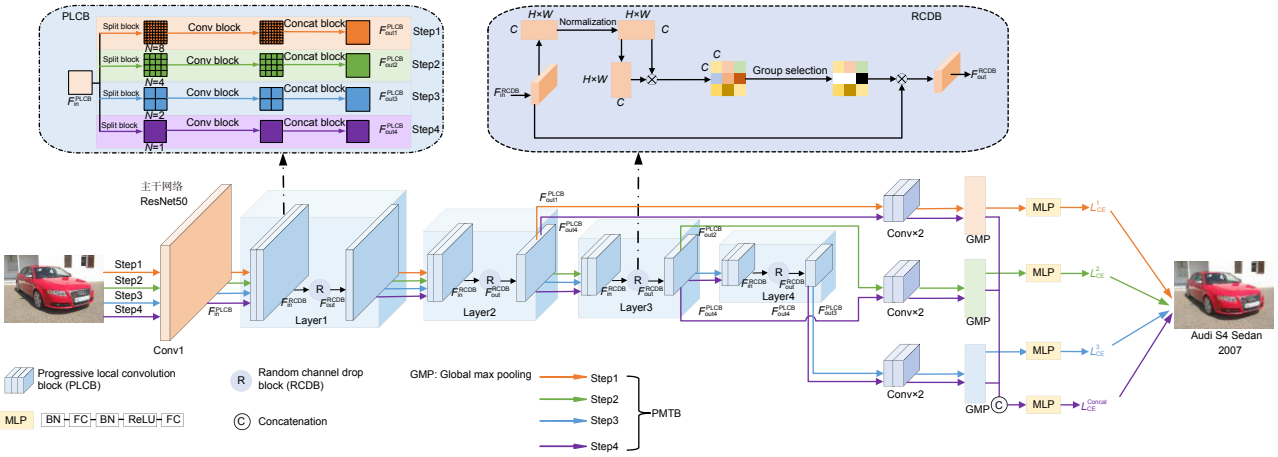
【通信作者】荆扬 (1996-)，男，山西运城人，硕士研究生，主要从事图像处理、细粒度识别等方面的研究。

E-mail：jingyang0525@xauat.edu.cn

扫描二维码，获取PDF全文

Progressive multi-granularity ResNet vehicle recognition network

Xu Shengjun[1,2], Jing Yang[1,2*], Li Haitao[3], Duan Zhongxing[1,2], Liu Fuyou[4], Li Minghai[1,2]



Overall structure of the proposed network

Overview: Model recognition aims to identify specific information such as the brand, model, and year of the vehicle, which can help verify the accuracy of tracking vehicle information. There are two research strategies for model recognition tasks. The strategy of strong supervision and learning involves utilizing image-level labeling information as well as additional bounding boxes in the model, component information, etc. Based on the strategy of weak supervision and learning, only the image-level label can be completely classified by fine particle size models. Most classification methods for weak supervision and learning adopt strategies such as attention mechanisms, dual-linear convolutional neural networks, and measurement learning. Pay more attention to the significant particle size of the vehicle's grid, tire tires, and other large granularity, and ignore the characteristics of small-size vehicle characteristics with distinguishing power such as car logo and door handles. Aiming at the difficulty of the vehicle due to the imaging differences such as posture and perspective, it is difficult to identify the model and propose a variety of multi-granular ResNet model recognition networks. First of all, using the ResNet network as the main network, propose a gradual multi-granular local convolution module to perform local convolution operations on vehicle images of different particle sizes, so that the network can be paid attention to the local characteristics of different particle-level vehicles when restructuring. Use the random channel discarding module to discard the multi-scale local feature map for random channel discarding, inhibit the network's attention to the characteristics of the vehicle's significant regional characteristics, and increase the attention of non-significant characteristics. Each training step is added to the classification loss. By dividing the network training process into different stages, the network can effectively integrate the multi-size features of the vehicle withdrawal, and guide the network extraction of multi-scale characteristics of vehicles with more discerning and diverse vehicles. The experimental results show that on the Stanford Cars dataset, the Compcars network dataset, and the model data set in the real scene, the accuracy of the network recognition accuracy has reached 95.7%, 98.8%, and 97.4%, respectively. Compared with the comparison network, the proposed network not only has the accuracy of recognition but also has better robustness. It has achieved very outstanding results in real scenes such as low light intensity and deformation of vehicles. The effectiveness of the model recognition on the road.

Xu S J, Jing Y, Li H T, et al. Progressive multi-granularity ResNet vehicle recognition network[J]. *Opto-Electron Eng*, 2023, **50**(7): 230052; DOI: 10.12086/oee.2023.230052

Foundation item: Project supported by National Natural Science Fundation of China (51678470, 61803293), Shaanxi Provincial Department of Education Special Fund (18JK0477, 2017JM6106), Shaanxi Province Natural Science Basic Research Fund (2020JM-472, 2020JM-473, 2019JQ-760), Basic Funding Project of Basic Research of Xi'an University of Architecture and Technology (JC1703, JC1706), Shaanxi Provincial Department of Science and Technology issued research projects (2021SF-429)

[1]College of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an, Shannxi 710055, China; [2]Xi'an Key Labratory of Building Manufactaring Intelligent & Automation Technology, Xi'an, Shannxi 710055, China; [3]Traffic Engineering Construction Bureau of Jiangsu Province, Nanjing, Jiangsu 210024, China; [4]CCCC Tunel Engineering Company Limited, Beijing 100024, China

* E-mail: jingyang0525@xauat.edu.cn