

基于 FR-ResNet 的车辆型号精细识别研究

余 焱^{1,2} 傅云翔¹ 杨昌东¹ 路 强^{1,2}

摘 要 车辆型号精细识别的关键是提取有区分性的细节特征. 以“特征重用”为核心, 以有效提取车辆图像细节特征并进行高效利用为目的, 提出了一种基于残差网络特征重用的深度卷积神经网络模型 FR-ResNet (Improved ResNet focusing on feature reuse). 该网络以 ResNet 残差结构为基础, 分别采用多尺度输入、低层特征在高层中重用和特征图权重学习策略来实现特征重用. 多尺度输入可以防止网络过深导致性能退化以及陷入局部最优; 对高层网络部分加以不同程度的特征重用, 可以加强特征传递, 高效利用特征并降低参数规模; 在中低层网络部分采用特征图权重学习策略, 可以有效抑制冗余特征的比重. 在公开车辆数据集 CompCars 和 StanfordCars 上进行实验, 并与其他网络模型进行比较, 实验结果表明 FR-ResNet 在车辆型号精细识别任务中对车辆姿态变化和复杂背景干扰等具有鲁棒性, 获得了较高的识别准确率.

关键词 车辆型号精细识别, 卷积神经网络, 残差结构, 特征重用

引用格式 余焱, 傅云翔, 杨昌东, 路强. 基于 FR-ResNet 的车辆型号精细识别研究. 自动化学报, 2021, 47(5): 1125–1136

DOI 10.16383/j.aas.c180539

Fine-Grained Car Model Recognition Based on FR-ResNet

YU Ye^{1,2} FU Yun-Xiang¹ YANG Chang-Dong¹ LU Qiang^{1,2}

Abstract The key of fine-grained car model recognition is to extract the discriminative feature details. Focusing on “feature reuse”, aiming at efficiently extracting and using the feature details from car images, a deep convolutional neural network model named FR-ResNet (improved ResNet focusing on feature reuse) is proposed. Based on the residual structure, FR-ResNet adopts the strategy of multi-scale input, reuse of low level feature in high level network, and weight learning of feature maps to realize feature reuse. Multi-scale input can prevent performance degradation caused by too deep network and fall into local optimum. Different degrees of feature reuse for high, medium and low level network can enhance the feature transfer, efficiently reuse the features and reduce the size of parameters. The strategy of feature map weight learning applied on middle and low level network can effectively suppress the proportion of redundant features. Experimental results carried out on the state-of-the-art public vehicle datasets CompCars and StanfordCars show that, compared with other network models, FR-ResNet can obtain high recognition accuracy in car model recognition task, and is robust in the shooting angle of vehicles and the change of background.

Key words Fine-grained classification of car models, convolutional neural network (CNN), residual structure, feature reuse

Citation Yu Ye, Fu Yun-Xiang, Yang Chang-Dong, Lu Qiang. Fine-grained car model recognition based on FR-ResNet. *Acta Automatica Sinica*, 2021, 47(5): 1125–1136

随着车辆的普及、交通运输技术以及智能交通系统的发展, 车辆识别作为计算机视觉、图像处理技术应用方面的重要研究课题, 具有较大的学术价值和前景. 无论是无人驾驶^[1]、停车场自动收费还

是公安部门对特定车辆的大范围检索^[2], 在不断上升的需求以及相关硬件设备性能升级的刺激下, 车辆识别在效率和精度上的要求也在不断提高. 套牌车、车牌污损、车标改装和违规喷漆等情况的存在使得单独的车牌、车标^[3]等识别技术难以奏效, 而车辆型号包含车辆品牌、车辆类型、所属系列和年代款式等信息, 其识别属于精细识别的范畴, 需要能识别出不同车辆之间的微小差异. 因此, 车辆型号精细识别将成为新一轮的研究聚焦点, 为智能交通部门及交管执法部门提供更好的解决方案.

目前, 车辆型号精细识别尚面临如下挑战:

- 1) 车型种类繁多. 目前国内已登记的车辆型号种类多达上万种, 即使是在一个中等大小的城市, 路面上常见的车辆型号也有上千种^[4].
- 2) 易受车辆姿态、拍摄环境的影响. 真实环境

收稿日期 2018-08-08 录用日期 2019-01-18
Manuscript received August 8, 2018; accepted January 18, 2019
国家自然科学基金 (61906061), 安徽省重点研究和开发计划项目 (201904d07020010) 资助
Supported by National Natural Science Foundation of China (61906061) and Provincial Key Research and Development Program of Anhui Province (201904d07020010)
本文责任编辑 白翔
Recommended by Associate Editor BAI Xiang
1. 大数据知识工程教育部重点实验室 (合肥工业大学) 合肥 230009
2. 合肥工业大学计算机与信息学院 合肥 230009
1. Key Laboratory of Knowledge Engineering with Big Data (Hefei University of Technology), Hefei 230009 2. School of Computer Science and Information Engineering, Hefei University of Technology, Hefei 230009

中,车辆的姿态是不固定的,所在场景、环境光照也各不相同.车辆姿态和环境的变化给车型识别增加了难度.

3) 车辆型号之间区分难度大.不同于 ImageNet^[5] 图像数据集中各分类对象种类间差异较大,车辆型号识别对象均为四轮汽车,即使目前数据集中的类别数量不足 1000,对应分类任务的难度仍不低于 ImageNet 分类任务.此外,部分车辆型号相互之间差别很小,这种细微的差距不仅存在于同品牌不同型号之间,不同品牌之间也有区分度很小的车型存在.如:图 1(a) 中的 2018 款的奥迪 A4L 和奥迪 A6L,它们的区别仅在于大灯和雾灯;图 1(b) 中的奇瑞 QQ3 和雪佛兰乐驰,它们的外形极其相似,区别仅在于散热格栅和雾灯部分.



(a) 奥迪 A4L 和奥迪 A6L
(a) Audi A4L and Audi A6L



(b) 奇瑞 QQ3 和雪佛兰乐驰
(b) Chery QQ3 and Chevrolet Spark

图 1 相似车型的例子

Fig. 1 Examples of similar vehicle models

传统采用尺度不变特征变换 (Scale-invariant feature transform, SIFT)^[6]、方向梯度直方图 (Histogram of oriented gradients, HOG)^[7] 等手工特征描述子的方法来实现特征提取,将提取到的特征利用分类器进行训练,以实现分类识别.这类方法易受光照、车辆姿态、噪声和背景等影响,鲁棒性不强.此外,无论是利用车前脸、车后脸还是它们的部分来进行车型识别,用手工特征描述子均难以描述,这是因为识别对象需要综合车灯、车栅格等诸多细节特征来进行判断,而手工描述子很难统一提取这些细节部位的特征.深度学习能够自动学习大量数据的特征^[8],借助于大数据,可以大大提升识别效率与精度,在人脸识别^[9]、行人检测^[10]、目标分类与发现^[11]、图像精细分类^[12] 等领域已经取得了很大的

成功.研究成果表明,深度卷积神经网络在车辆型号识别领域的应用可以获得很好的识别效果.

虽然目前已有基于深度学习的商用车型识别系统投入实际使用,但其针对的识别对象为卡口监控中获取的车辆正脸图像,姿态变化很小.实际智能交通应用中,涉及的车辆图像包含从多个角度拍摄的图像,因此,如何基于多姿态车辆图像进行车辆精细型号的识别,具有十分重要的研究价值和现实意义.

针对多姿态车辆型号的精细识别,本文提出一种基于残差网络特征重用的深度卷积神经网络模型 FR-ResNet (Improved ResNet focusing on feature reuse),该网络具有以下特点:

1) 网络采用残差结构作为基础,在加深网络层数的同时,不必担心梯度爆炸问题和随着层数增加发生准确率下降的网络退化问题.

2) 添加多尺度信息输入,加强图像多尺度特征融合,防止网络陷入局部最优.

3) 对不同层网络特征采取不同程度的特征重用,将上一层特征与当前层特征进行融合,以促进特征流动,提高其利用率,有效缩减所需参数数量.

4) 在底层网络中采用特征图权重学习策略,在图像处理初期对特征通道的重要程度进行排序,使有效特征得到更大激励并传递下去.

采用目前流行的大型车辆图像数据集 CompCars 和 StanfordCars 为测试数据,在此之上进行一系列实验.实验结果表明 FR-ResNet 在车辆型号识别方面性能优于其他一些经典的卷积神经网络 (Convolutional neural network, CNN) 模型.例如在 CompCars 数据集上, VGG16^[13] 获得了 92.4% 的准确率, ResNet^[14] 获得了 93.7% 的准确率,而本文提出的 FR-ResNet 的准确率达到 95.1%.

1 相关工作

“车型识别”包含两个含义,即“车辆类型识别”和“车辆型号识别”.早期“车型识别”主要指“车辆类型识别”,即识别车辆是小汽车、卡车还是公交车等.如:文献 [15] 中使用稀疏拉普拉斯滤波器来学习大量未标注数据,使用半监督的卷积神经网络自动为分类任务学习有利特征,在复杂场景中对巴士、小型巴士、小型货车、乘用车、轿车和卡车共 6 种车辆类型识别效果良好.车辆型号由于种类繁多、类间区分度小等原因,其识别工作难度远远大于车辆类型的识别,属于精细分类问题.

目前车辆型号精细识别方面的研究工作还较少,研究方法主要可以分为三类:基于“传统特征描述子 + 分类器”的方法、基于 3D 模型的方法以及基于深度学习的方法.

基于“传统特征描述子 + 分类器”的方法是较为传统的图像识别方法. 一般使用人工描述子提取图像特征, 再用分类器训练以实现分类. Hsieh 等^[16]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 (Speeded up robust feature) 描述子提取特征, 并在每个网格块上使用支持向量机 (Support vector machine, SVM) 训练弱分类器, 基于分类结果的组合进行最终车辆型号的识别. Liao 等^[17]提出了一种基于车辆部件的分类方法, 采用强监督 DPM (Deformable parts model) 来引入语义层次结构进行语义分割, 基于部件的外观和语义来识别车辆. Biglari 等^[18]则通过捕捉不同种车辆整体外观以及各部件间的特征差异, 通过 SVM 训练获得相应的车型特征模板. 文献 [19] 提出一种基于部件矫正的光度特征提取算法, 增强了不同光照强度下摄像头拍摄照片识别的稳定性. 但是由于人工设计特征的局限性, 研究处理的图像多为车辆正面图像, 姿态单一.

为了能更好地应对图像视角的变化, 研究者们提出了基于 3D 模型的方法, 3D 模型能体现局部特征和模型整体之间的空间关系. 文献 [20] 中提出一种 3D 对象建模和精细分类任务相结合的方法, 详细的 3D 表示相对 2D 有更多的特征信息, 改善了精细分类的性能. Krause 等^[21]认为单纯的平面图形限制了视角, 提出一种从 2D 图形提取 3D 模型的方法, 通过 CAD 建模和特征贴片, 最终形成 3D 训练特征, 实验结果优于此前的 2D 方法.

自从 Krizhevsky 等的论文^[12]引起深度学习研究热潮以来, 神经网络影响巨大, 推动了多个领域研究工作的长足发展. 在车辆型号识别方面, 深度学习也起到了重要的作用. 不少研究者借助于卷积神经网络进行车辆型号分类的研究.

基于深度学习的识别方法主要通过建立端到端的卷积神经网络模型来完成识别分类任务. 目前典型的卷积神经网络模型 AlexNet, VGGNet, GoogLeNet, ResNet 等均能用于车辆型号识别, 但由于识别任务的特殊性, 这些网络在识别性能上并不十分出色. 研究者们试图在经典网络的基础上进行改进, 以使卷积神经网络模型更适用于解决车辆型号精细识别问题. Sochor 等^[22]将多角度的车辆图像通过 3D bounding box 进行边界限定, 并基于 3D 边界框对 2D 图像进行 3D 展开以及栅格化. 把展开后的图像作为深度学习的输入, 两者结合效果提升明显. 文献 [23] 在文献 [24] 提出的 TCNN (Tiled CNN) 的基础上改变了 CNN 的权重共享方案, 提出一种局部平铺卷积神经网络模型 LTCNN (Local CNN) 用于车型识别, 该网络具有平移、旋转和尺度不变性. 文献 [25] 首先借助 RCNN (Regions with

CNN) 从复杂背景图像中识别出车辆, 然后利用联合贝叶斯网络计算类间和类内相似度, 以获得车辆最大概率所属型号类别. 文献 [26] 提出一种特征融合的卷积神经网络模型, 把车正脸图像分上下两部分, 并行提取特征并将其多维度融合后用来实现车型的精细分类. 文献 [27] 尝试使用定位、识别两部分网络, 并采用多任务机制, 将车辆定位与显著性检测同时进行, 能够处理并识别背景杂乱无章的车辆图像. 文献 [28] 结合车辆部件检测的方法, 通过 CNN 网络获得图像整体和局部部件的特征, 通过 SVM 进行分类识别.

综上所述, 车辆型号精细识别研究目前的突破点和热点在于深度学习的应用. 识别对象类间差距小、种类繁多是研究的难点. 为有效提取车辆图像特征并进行高效利用, 提出了“特征重用”的思路, 围绕这一思路对 ResNet 进行改进, 提出了 FR-ResNet 模型, 并基于该模型, 实现了车辆型号的精细分类.

2 算法描述

2.1 残差结构

在目标识别领域, 卷积神经网络一直向着更深的模型结构发展, 识别效果得到不断提升. 更深更宽的网络模型具有更多的网络参数, 能够更好地学习到图像中的细节特征, 然而深层网络难以训练、更深网络反而训练误差加大、网络反向传播的参数相关性降低等问题日益明显. 为此, He 等^[14]提出了残差网络, 即通过残差表示和快速链接的方式来解决上述问题.

如图 2 所示, $G(x)$ 表示待拟合对象, x 为输入数据, 则残差函数 $R(x)$ 可以表示为 $G(x) - x$, 最终

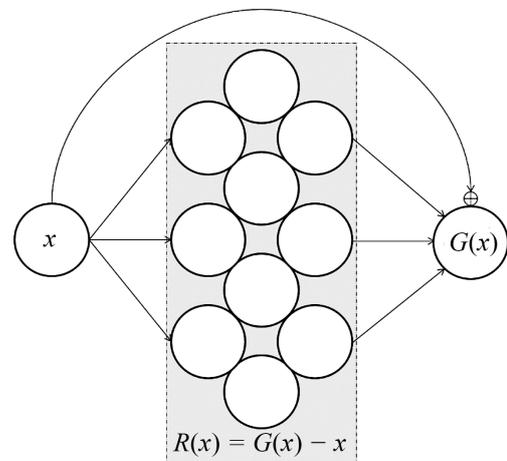


图 2 残差函数拟合关系

Fig. 2 Fitting relationship of residual function

拟合函数 $G(x) = R(x) + x$. 通过卷积神经网络的反向传播机制, 学习残差函数 $R(x)$ 的参数来拟合 $G(x) - x$, 以达到进一步拟合 $G(x)$ 的效果. 由于残差结构可以在增加网络层数的同时有效阻止网络退化, 因此, FR-ResNet 采用残差结构作为网络的主体结构, 基于残差结构的串联组合, 结合特征重用、权重学习和多尺度输入的策略, 构建适用于车辆型号精细分类的卷积神经网络模型. 在 FR-ResNet 中, 单个残差结构的组成如图 3 所示, 由 1×1 全通道卷积与 3×3 分组卷积的卷积层组合来实现.

2.2 网络主体结构

车辆型号精细识别网络 FR-ResNet 的主要模型结构如图 4 所示. 该模型由 3 个部分组成, 按照网络中生成特征图的大小分为低层网络、中层网络和高层网络. 低层网络中的特征图尺寸范围为 56×56 像素到 224×224 像素, 中层网络的特征图尺寸范围为 14×14 像素到 56×56 像素, 高层网络的特征图尺寸为 7×7 像素. 每层网络部分都包含一个残差模块, 由多个第 2.1 节所述的残差结构串联组合而成. 其中, 低层网络部分组成残差模块的残差结构为 6 层, 中层网络部分的残差结构高达 20 层, 高层网络部分的残差结构为 3 层.

FR-ResNet 模型的设计动机描述如下: 为防止陷入局部最优解并充分利用原图像所含特征, 采用多尺度输入的方法来增加数据输入; 为提高更低层结构中的特征利用效率, 采用特征重用的方法来避免有效特征的流失; 为合理分配不同特征图的比重, 采用特征图权重学习的方法来加强网络中有效特征的流动.

2.3 多尺度输入

在目标检测和识别中, 图像金字塔是经常使用的结构, 即把图像表示为一系列分辨率逐渐降低的图像集合, 这是一种多尺度的表达方式, 可用于数据

的扩增. 在深度学习中, 多尺度信息的输入可以防止设计的卷积神经网络陷入局部最优解, 促进网络参数的更新. 因此本文在网络中增加了多尺度信息输入的方法.

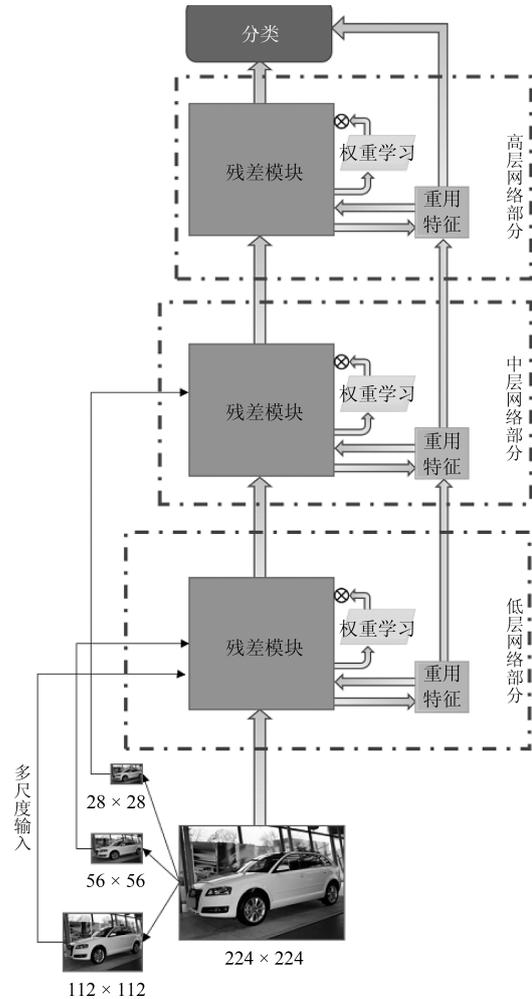


图 4 网络结构示意图

Fig. 4 Diagram of network structure

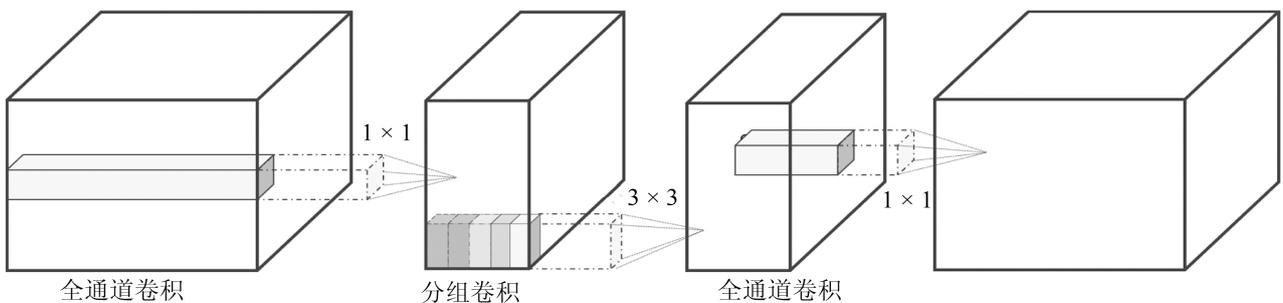


图 3 FR-ResNet 中的残差结构

Fig. 3 Residual structure in FR-ResNet

如图 4 中最下端数据输入部分所示, 多尺度输入需要在网络输入端对图像矩阵进行多尺度降采样. 为了与网络中特征图尺寸匹配, 降采样后的尺寸分别为 112×112 像素、 56×56 像素和 28×28 像素, 将结果分别送到生成同样尺寸特征图的中间层进行通道组合. 多尺度特征信息的融合可以提高数据的特征利用率. 在训练过程中, 使用 dropout^[29] 技术进行随机丢弃, 以产生网络局部震荡, 促进部分网络参数的更新以防止陷入局部最优解, 同时还可以避免网络过拟合. 多尺度输入在网络输入前就准备好了多尺度信息, 而不是在网络内部生成, 因此在网络训练中不需要额外增加参数, 维持了网络的原有计算成本.

2.4 特征重用

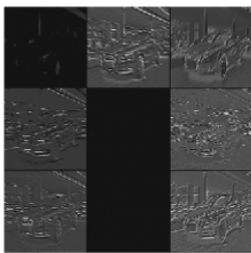
残差结构的存在允许设计出一个高深度的网络而不用担心其训练的退化问题, 然而在训练过程中, 由于大多数参数的作用范围很小, 且每层检测到的特征直接作为下一层的输入, 在更深层的信息流动过程中该特征的影响极小, 很多信息会在特征传递过程中丢失, 而每一层只能从上一层获得数据, 丢失的信息无法补充.

不同于普通分类任务, 车辆型号识别属于精细分类问题, 所需要的特征量更多也更细致. 各种型号

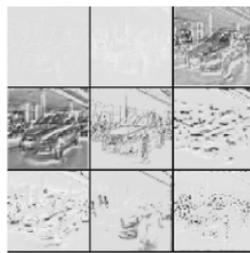
的车辆, 其车辆轮廓、车窗、车灯、车栅格、车门、倒车镜等部件的形状、位置均存在不同. 这些不同的特征中既包含宏观的外观信息, 也包含细节的纹理信息, 还包含空间位置信息. 本文使用卷积神经网络模型来实现识别任务, 数据信息从网络输入端输入后, 每层网络所提取到的特征各不相同. 对每层所获得的特征图进行可视化, 可以看出, 随着数据在网络中的流动, 所得到的特征信息愈来愈抽象. 我们通过一个例子来直观展现这一现象. 图 5 为车辆图像在深度卷积神经网络不同层次所得特征图的部分可视化结果. 输入为三通道 RGB 图像, 图中颜色的深浅描述网络中的神经单元在不同区域的激活程度. 从中可以看出, 卷积网络在浅层阶段提取的信息更简单, 如边缘信息, 从图中仍能看出这是一辆车. 而在深层特征图中提取到的不再是低层特征, 开始包含纹理等多样化的高层特征, 且感受野范围也更大.

每层均基于上一层的输出进行学习, 因此越深层的特征越抽象. 但也因为如此, 在层层传递的过程中, 浅层的一些信息可能会连同无用的背景信息一起被过滤掉.

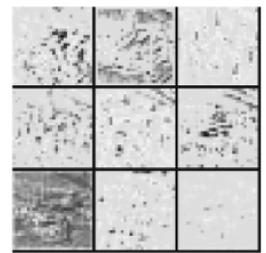
为防止在最终决策中错过最初的有效信息, 即避免特征信息的浪费, 提出了一种特征重用的方法, 对不同层次模块的特征进行不同程度的重用, 弥补上层特征的丢失, 以达到使最终生成的特征更加全



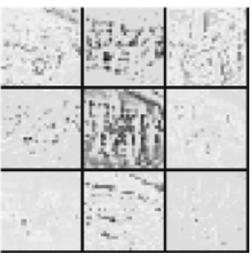
(a) 低层网络部分第一个卷积层
(a) First conv layer at the low-level



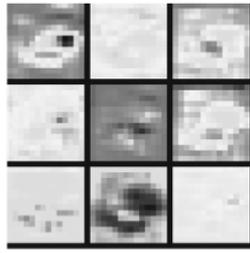
(b) 低层网络部分第二个卷积层
(b) Second conv layer at the low-level



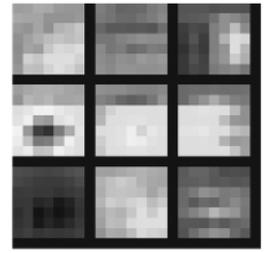
(c) 低层网络部分最后一个卷积层
(c) Last conv layer at the low-level



(d) 中层网络部分第一个卷积层
(d) First conv layer at the middle-level



(e) 中层网络部分最后一个卷积层
(e) Last conv layer at the middle-level



(f) 高层网络部分最后一个池化层
(f) Last pool layer at the high-level

图 5 特征图可视化

Fig. 5 Visualization of feature maps

面的目标.

特征重用的主要结构为

$$x_n = \text{Concat} [P_n(x_{n-1}), R_n(x_{n-1})] \quad (1)$$

其中, x_n 为第 n 层的输出, P_n 表示重用算法, $P_n(x_{n-1})$ 表示在第 n 层输出中选取第 $n-1$ 层前 $1/P$ 个通道的特征图作为重用特征图, R_n 为主体网络的残差运算, $R_n(x_{n-1})$ 表示将第 $n-1$ 层输入图 3 所示残差结构后输出. Concat 操作将两组特征图在通道上进行合并.

如图 6 所示, 假设特征图的大小用 $N \times C \times H \times W$

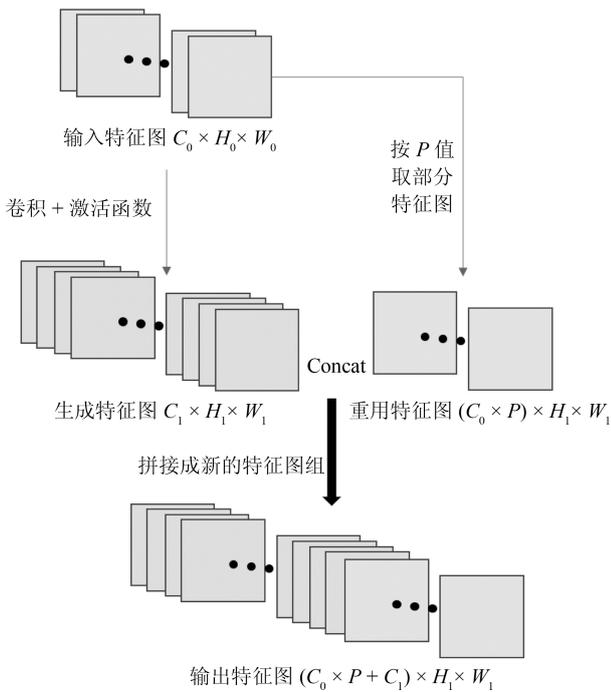


图 6 特征重用过程

Fig. 6 The process of feature reuse

W 表示, 其中, N 为 $batch_size$ 大小, C 为通道数, H 为特征图高度, W 为特征图宽度. 若当前结构的输入特征图尺寸为 $N \times C_0 \times H_0 \times W_0$, 生成的特征图大小为 $N \times C_1 \times H_1 \times W_1$, 重用部分的比例为 P . 特征重用应用于网络中所有的残差结构 (如图 3 所示), 即: 上一个残差结构的输出中的部分信息会作为特征重用信息, 与当前残差结构的输出结合在一起, 作为最终输出的特征图组. 其中, 特征重用参数

P 的取值选择问题将在第 3.5 节中详细描述. 重用的特征图根据生成特征图的比例进行适当调整, 重用特征图为 $N \times (C_0 \times P) \times H_1 \times W_1$. 则最终输出特征图为 $N \times (C_0 \times P + C_1) \times H_1 \times W_1$.

2.5 特征图权重学习

卷积神经网络的核心是卷积层, 卷积核在一组特征图上进行卷积操作生成新的特征图, 多个卷积核生成的特征图共同构成卷积层的输出. 针对生成的特征图, 一般深度学习网络中在进行后续操作时, 对每张特征图是同等对待的. 然而, 实际上每个特征图所包含的有价值信息不同, 对车型识别任务的贡献也各不相同. 如图 5 所示, 同一特征图组中每张特征图所包含的有效信息各不相同, 有的包含了很重要的有效信息, 有的包含的有效信息内容较少, 因此尽可能多地保留这些有效信息, 加强有效特征信息在网络中的流动很有必要.

为实现增强有效特征信息、抑制噪声的目的, 本文采用一种特征图权重学习的策略, 通过网络学习的方法, 为同一层的每个特征图分配新的权重. 具体实现如图 7 所示, 将尺寸为 $C \times H \times W$ 的特征图组输入一个双通道的降维通道, 降维通道的池化层选择策略与对比实验将在第 3.6 节详细说明. 本文中的通道由局部最大池化与全局平均池化层组成.

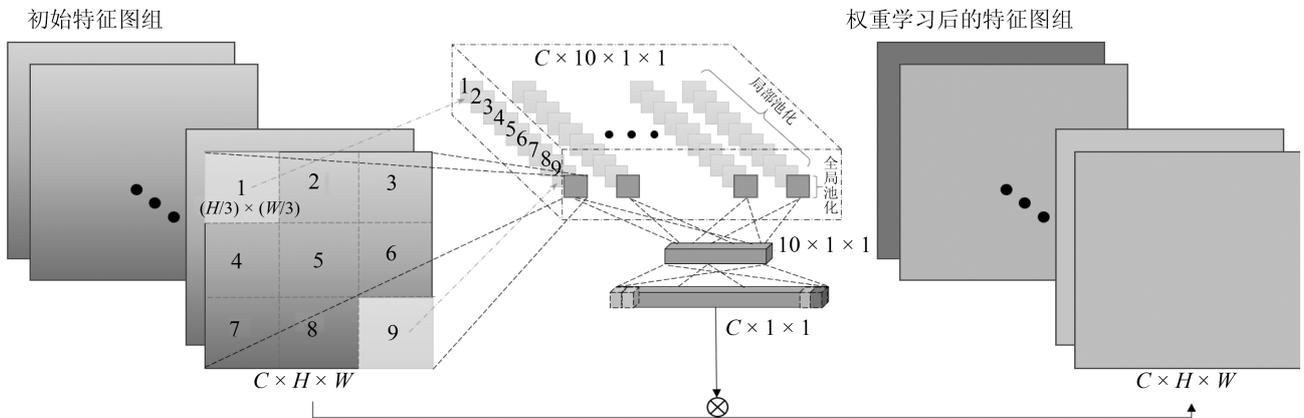


图 7 特征图权重学习

Fig. 7 Weight learning based on feature maps

其中局部最大值池化是将特征图划分为 3×3 共 9 个局部区域, 分别进行最大值池化. 通道的输出被压缩后排列为一维数据, 在对其进行两次全连接操作后, 利用 Sigmoid 函数生成一维权重参数 $C \times 1 \times 1$, 参数深度等于特征图的个数. 最后将权重与原特征图矩阵相乘获得新的特征图组. 由于高层网络中特征图用于最终分类, 其尺寸小且数量多. 图 7 所示的先降维再升维的权重学习方法无法在过小的尺寸上通过局部和全局池化获得更多的有效信息, 提升效果不显著. 反而会因为特征图数量多, 导致全连接参数大大增加, 额外增加了计算时间, 所以只将该方法应用在低层和中层网络中.

特征图权重学习策略使用局部和全局池化来提取各特征图矩阵信息, 可以在节省参数个数的同时, 有效提取特征图各自的特征以及整个特征图组的空间信息, 然后通过学习获得每个特征图的权重, 将权重传入下一层的特征图组中, 以提高有效特征信息的比重, 从而增强有效信息的流动.

3 实验结果与分析

3.1 数据集

车辆型号精细识别研究所使用的数据集分为两类: 监控数据集和网络数据集. 监控数据集中的图像主要来源于卡口监控系统, 网络数据集中的图像主要来源于网络. 由于卡口监控设备的固定性, 卡口抓拍图像中的车辆姿态固定, 一定程度上降低了其分类的难度. 网络数据集中的图像所在场景多变、姿态各异, 且分辨率也不固定, 因此, 基于网络数据集的车型识别更具有挑战性.

采用网络数据集中具有代表性的 StanfordCars 和 CompCars 网络数据集为实验数据集, 开展实验研究 (数据集中部分图像如图 8 所示). StanfordCars 数据集共含车型 196 类, 总图像数量为 16 185 幅. 数据集标签包括车型种类及车辆在图像中的定位信息. CompCars 数据集包含监控数据和网络数据两个部分, 其网络数据集共记录了 431 种车型、5 种车辆拍摄姿态, 总图像为 52 083 幅, 包含车辆位置标定以及车部件等细节信息.

3.2 实验环境与设置

实验的硬件环境如下: CPU: Intel Core i7-7700K CPU @ 4.20 GHz \times 8; 内存: 16 GB; 显卡: Nvidia GTX1080Ti; 显存: 12 GB.

实验所有模型在 Ubuntu16.04 环境下基于开源框架 Caffe^[30] 实现, CUDA 版本为 9.0.

在识别速度、所需硬件条件差别不大的情况下, 识别准确率是衡量识别算法性能的重要指标, 准确率的计算与对比也很简洁明了. 为易于本文实验结

果与其他算法结果的比较, 使用常用的准确率计算公式, 即

$$\text{准确率} = \frac{\text{正确预测样本数}}{\text{预测样本总数}} \times 100\% \quad (2)$$



(a) StanfordCars 数据集中的样本
(a) Samples from StanfordCars dataset



(b) CompCars 数据集中的样本
(b) Samples from CompCars dataset

图 8 数据集中的样本

Fig. 8 Samples from datasets

FR-ResNet 的实验准确率基于实验环境进行计算获得, 本文实验部分的其他对比实验, 其准确率数据来源为: 1) 优先采用对应论文提供的实验数据; 2) 经典网络模型采用与 FR-ResNet 相同的训练设置以及数据预处理方式, 通过官方提供的代码复现.

网络训练采用 224×224 像素的数据输入尺寸, 以与其他网络原有的输入保持一致. 每幅输入图像的预处理操作使用文献 [12] 中提到的方法, 先将图像大小归一化到 256×256 像素, 裁剪 4 个角外加中心部分获得 5 幅 224×224 像素图像, 并各自进行镜像操作, 共生成 10 幅训练图像, 再从中减去训

练集图像均值,即为训练输入数据.在测试时,对图像进行减均值操作时使用训练集均值,对10幅扩展出的图像求平均预测值.网络使用随机梯度下降法(Stochastic gradient descent, SGD)更新模型权重,初始学习率为0.001,每10万次迭代降低10倍学习率.整个训练过程共迭代35万次.

3.3 在 StanfordCars 数据集上的实验

基于 StanfordCars 数据集,使用其提供的所有图像数据用于实验.为了与其他研究成果进行更全面的比较,数据集输入采用原图,即无 BBox (bounding box) 和带 BBox 两种方式.

实验结果如表1所示,在使用 BBox 的情况下, BB-3D-G (3D BubbleBank which pooling bubble responses globally) 为 StanfordCars 数据集发布时作者实验的结果,作为实验对比的参照,识别准确率为 67.6%. 在深度学习还未广泛应用时, LLC (Locality-constrained linear coding)^[31] 在编码方式上使用局部限制线性编码方式,其准确率略有提高,为 69.5%. ELLF (Ensemble of localized learned features)^[32] 通过重点学习车辆部件等细节特征,使识别率达到 73.9%. VGGNet 在多个迁移学习任务表现优秀,应用在车辆型号识别领域也有着一定的效果,在不使用 BBox 的情况下,识别准确率为 75.6%. 2015 年, ResNet 的出现加深了卷积神经网络的层数, ResNet-101 就能达到 85.8% 的准确率,体现了网络深度在车型识别方面的优势. FCANs (Fully convolutional attention networks)^[33] 依靠在视觉提取以及部件注意机制方面的创新,将识别率提升到 89.1%,如果使用 BBox 准确率还能再提高两个百分点.文献[34]中提出具有集合约束的 triplets,获得了 92.5% 的准确率,

表1 在 StanfordCars 数据集上的实验结果比较 (%)

Table 1 Comparison of classification results on the StanfordCars dataset (%)

模型方法	准确率(有 BBox)	准确率(无 BBox)
BB-3D-G ^[21]	67.6	-
LLC ^[31]	69.5	-
ELLF ^[32]	73.9	-
VGGNet ^[13]	-	75.6
ResNet	85.8	74.8
FCANs ^[33]	91.3	89.1
Wang 等 ^[34]	92.5	-
Krause 等 ^[35]	92.8	-
FR-ResNet	93.1	90.6

Krause 等在文献[35]中使用 BBox, 准确率能达到 92.8%. 我们的网络在使用 BBox 的情况下能达到 93.1% 的准确率,即使不使用 BBox, 准确率也有 90.6%, 比 ResNet 使用 BBox 的识别效果还要好.

3.4 在 CompCars 数据集上的实验

CompCars 数据集包括 2 类,即卡口监控数据集和网络数据集.卡口监控数据集中的车辆图像共 44 481 幅、281 类,其中训练样本 31 148 幅、测试样本 1 333 幅.在此数据集上进行实验,多个经典模型均能获得 97% 以上的 Top-1 准确率,其中 AlexNet, GoogLeNet 和 ResNet 的 Top-1 准确率分别为 97.8%, 98.5% 和 98.8%, 本文提出的 FR-ResNet, 其 Top-1 准确率为 99.2%. 而上述模型的 Top-5 准确率都能达到 99.5% 以上.虽然卡口图像分辨率低,但是图像上的车辆姿态固定,因而降低了分类难度,各分类模型的识别准确率都很高.为有效体现本文方法的优势,采用 CompCars 网络数据集来进行实验的详细分析.

CompCars 网络数据集提供的车辆图像总数高达 143 060 幅,标注部分为其中的 36 456 幅.以标注的 36 456 幅为训练样本,提取 15 627 幅为测试样本,在此基础上进行实验.为体现 FR-ResNet 网络在该数据集上的良好表现,使用近些年多个经典的卷积神经网络模型以及其他研究者的研究成果与本文的网络进行比较,实验结果如表 2 所示.

表2 在 CompCars 数据集上的实验结果比较 (%)

Table 2 Comparison of classification results on the CompCars dataset (%)

模型方法	Top-1 准确率	Top-5 准确率
Yang 等 ^[36]	76.7	91.7
AlexNet	81.9	90.4
BoxCars ^[22]	84.8	95.4
GoogLeNet ^[37]	91.2	98.1
ResNet	92.3	98.4
DenseNet ^[38]	93.1	98.6
Hu 等 ^[27]	94.3	98.9
FR-ResNet	95.3	99.1

CompCars 网络数据集中的车辆具有各种不同的姿态,该数据集的发布者针对其所有姿态进行实验,获得了 Top-1 为 76.7%, Top-5 为 91.7% 的实验结果.早年的 AlexNet 由于自身网络结构的局限性,在精细识别任务上较为乏力, Top-1 与 Top-5 准确率仅为 81.9% 和 90.4%. 文献[22]通过建立 3D 包围盒获取车辆额外 3D 信息,取得 Top-1

为 84.8%, Top-5 为 95.4% 的准确率. GoogLeNet 凭借其网络深度和宽度, 将 Top-1 提高到 91.2%, Top-5 提高到 98.1%. ResNet 效果更好, Top-1 为 92.3%, Top-5 为 98.4%. 文献 [27] 采用先定位后识别的方法, Top-1 准确率为 94.3%, Top-5 准确率为 98.9%. 总体来看, Top-1 准确率的提升远大于 Top-5, 可见准确率的提升更多依靠的是对类间差小的车型的正确识别. 而本文提出的 FR-ResNet 模型仅使用分类标签, Top-1 和 Top-5 的准确率分别达到了 95.3% 和 99.1%, 平均每种类别的错误张数不足 0.4 张.

对分类错误的图像进行分析, 并总结错误的原因, 可能有: 1) 复杂背景和环境光照影响了车辆的正确识别 (如图 9 (a) 所示); 2) 车辆本身发生变化, 如图 9 (b) 所示, 奥迪车车门敞开, 使车身部分的有效特征发生了变化; 3) 车辆在原图中占比过小, 如图 9 (c) 所示; 4) 受拍摄角度的影响, 如图 9 (d) 所示, 车辆图像是自顶向下拍摄的, 而在训练集中缺少相同拍摄角度的样本.

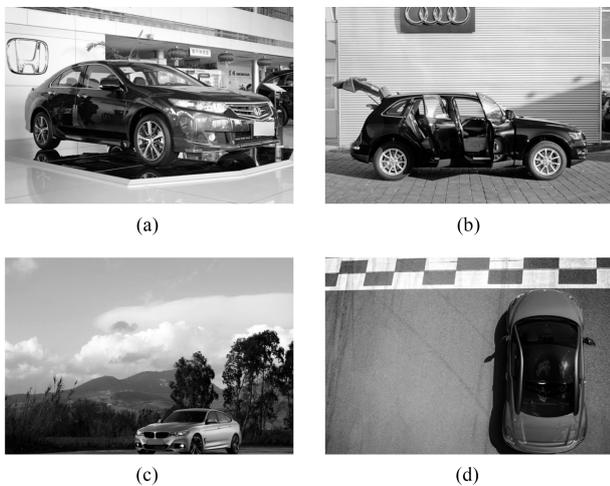


图 9 识别错误的样本

Fig. 9 Samples of error recognition

为体现网络在特征提取方面的优势, 并验证网络的鲁棒性, 减少训练样本的数量, 以验证在少量样本情况下网络仍具有很好的性能. 按照数据集官方早期提供的标注信息, 其中训练样本数量不足目前标注训练样本数量的一半, 我们提取了数据集中旧版标注的 16 016 幅为训练样本, 14 939 幅为测试样本, 生成少量样本数据集 A, 在此基础上继续进行实验. 本文提出的 FR-ResNet 在此数据集上的表现, 以及与表 2 中效果较好的几种经典网络结构的对比结果如表 3 所示. 在没有足量训练数据的情况下, 各网络的识别准确率均有所降低. GoogLeNet 的准确率 Top-1 只有 65.8%, Top-5 为 87.9%. 更深的

ResNet 仅有 78.3% 的 Top-1 的识别准确率, Top-5 准确率也只有 93.5%, 单纯的残差结构效果并不好. DenseNet 通过加强信息传递 Top-1 识别准确率可以上升到 90.6%, Top-5 准确率也提高了接近 5 个百分点, 说明特征利用率的提升对车辆型号精细识别有效. 本文中提出的 FR-ResNet 更加专注特征信息的有效利用, 其准确率得到了更进一步的提升, Top-1 识别率达到了 92.5%, Top-5 识别率达到了 98.4%. 为了进一步检验网络性能, 我们还尝试使用更小数量的训练样本进行实验. 在 16 016 幅训练样本的基础上每类再次减少 1/3, 形成少量样本数据集 B. 在 Top-1 识别率上, GooLeNet 仅为 52.3%, ResNet, DenseNet 分别为 69.7% 和 81.5%, 而我们的 FR-ResNet 达到 85.2%, 识别效果更好. 再次减少训练样本数量, 取 16 016 幅训练样本的 1/2 时, 训练过程困难, 实验结果不佳, 每个网络的准确率均不超过 60%. 造成这种结果的原因是, 原数据集每类样本的数量不均匀, 当训练数据减半时, 多数种类样本不足 20 幅, 过少的训练样本使得深度学习网络无法完成很好的收敛.

表 3 在 CompCars 少量样本数据集上的实验结果比较 (%)

Table 3 Comparison of classification results on the small training samples from CompCars dataset (%)

模型方法	少量样本集 A		少量样本集 B	
	Top-1 准确率	Top-5 准确率	Top-1 准确率	Top-5 准确率
GoogLeNet	65.8	87.9	52.3	76.7
ResNet	78.3	93.5	69.7	82.3
DenseNet	90.6	98.0	81.5	90.1
FR-ResNet	92.5	98.4	85.2	93.8

3.5 特征重用比例

FR-ResNet 在采用特征重用策略提升网络性能的同时, 也增大了网络的规模, 特征重用的比例 P 随着网络深度、特征图数量的变化需要有所调整, 以获取最佳实验效果.

本文使用 Caffe 框架中的 Slice 层来实现通道选择的功能. Slice 层的作用是通过设置多个分割点 $slice_point$ 来分割输入的特征图, 其中参数 $axis$ 可以设置特征图分割的维度. 由于特征图中各通道在初始化阶段等价, 故选取需要设置参数最少的前 $1/P$ 个通道进行重用. 本文中设置 $axis = 1$, 将特征图按通道分割, 使用单个分割点, 使得分割点左侧的通道数与总通道数比值为 $1/P$, 以达到选取 $1/P$ 个通道的目的.

根据深度及特征图数量将网络分成 4 个不同的阶段, 依次进行 P 值的调优实验, 当取值跨度较大时, 可通过二分法确定选值. 第 1 阶段为低层网络部分, 特征图组中的特征图数量为 256, 分别使用 1/64, 1/16, 1/8 的 P 值比例进行实验, 结果如表 4 所示, 故选择 1/16 作为该阶段 P 值. 第 2 阶段同为低层网络部分, 特征图组中的特征图数量为 512, 根据实验结果选择 1/16 作为该阶段 P 值. 第 3 阶段为中层网络部分, 特征图组中的特征图数量为 1024, 使用 1/64, 1/16, 1/8 的 P 值比例进行实验, 实验结果表明 P 值为 1/64 与 1/16 时提升效果相近但重用特征图数量相差较多, 故使用二分法选择更优的 P 值. 最终确定该阶段 P 值为 1/32. 第 4 阶段特征图组中的特征图数量为 2048, 实验选取 P 值为 1/16.

3.6 特征图权重学习降维通道池化层选择策略

FR-ResNet 采用特征图权重学习来增强有效信息比重, 如果直接用全连接层提取特征图信息, 以用于权重学习, 将导致参数量剧增, 不利于深层网络的训练. 添加池化层是常用的降维方法, 能在不额外增加参数的同时提取到有效特征. 根据所提取特征的不同和池化方法的不同, 可以采取全局平均、全局最大值、局部平均、局部最大值 4 种单独的池化方案, 以及其两两组合的方案.

基于上述方案分别进行实验, 实验结果如表 5 所示, 从中可以看出单独的池化策略中, 全局平均池化的效果最好; 而在池化的组合方案中, “全局最大值 + 局部最大值” 的提升效果最弱, 可能原因是局部最大值已经包含了全局最大值; “全局平均池化 + 局部最大值池化” 效果最好, 因此本文选择了这种组合的池化策略.

3.7 各改进策略的效果分析

FR-ResNet 以特征的高效利用为目标, 分别采用了三种改进策略来完善网络模型, 即: 多尺度输入、特征重用和特征图权重学习. 为了证明这些方法的有效性, 使用控制变量的研究方法在 CompCars 数据集上进行对比实验. 实验结果如图 10 所示, 从中不难看出, 三种结构均能提高网络识别准确率. 其中, 特征重用的效果最为明显, 其识别率的提升最

高, Top-1 和 Top-5 的识别率分别提升了 2.7% 和 0.6%. 由此可见, 数据在网络传播中会丢失部分有效特征, 且较低层特征在较高层网络中对识别任务仍然具有贡献. 特征图权重学习策略位居第二, 其 Top-1 和 Top-5 识别率分别提升了 1.5% 和 0.3%, 说明了增强有效信息流动的可行性. 多尺度输入策略对网络影响较小, 但识别率仍有一定的提升, 说明该方法仍然存在有效性. 提升的结果符合该方法预期的反馈效果, 即其在数据扩增和避免局部最优方面的作用.

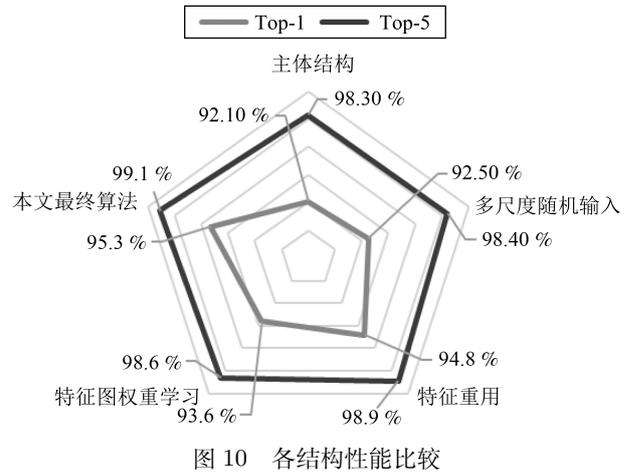


图 10 各结构性能比较
Fig. 10 Comparison of performances of all structures

4 结论

本文针对车辆型号精细识别问题进行研究, 提出了以残差结构为网络主体结构, 以特征重用为主要思想的方法, 并基于此设计了深度卷积神经网络模型 FR-ResNet. FR-ResNet 采用了多尺度数据输入、低层特征在高层中重用和特征图权重学习三大策略, 有效阻止了网络训练的退化以及陷入局部最优解, 提高了对网络中有效特征的利用率, 减轻了冗余信息的干扰. 在 CompCars 和 StanfordCars 两大公开数据集上的测试结果表明 FR-ResNet 在车辆型号精细识别上具有较高的识别精度, 高于其他的一些网络模型. 此外, FR-ResNet 在车辆姿态变化、复杂背景干扰、训练样本减少时, 具有一定的鲁棒性.

表 4 特征重用比例 P 值对准确率的影响

Table 4 Effect of feature reuse ratio P on recognition accuracy

P 值	第 1 阶段			第 2 阶段			第 3 阶段			第 4 阶段				
	1/64	1/16	1/8	1/64	1/16	1/8	1/64	1/32	3/64	1/16	1/8	1/64	1/16	1/8
准确率 (%)	93.6	93.7 %	93.4	93.6	94.1 %	93.9	94.3	94.6 %	94.2	94.2	94.0	94.6	94.8 %	94.5 %

表 5 权重学习中池化策略的对比 (%)

Table 5 Comparison results of pooling strategies in weight learning (%)

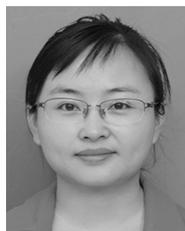
池化选择策略	Top-1 准确率
全局平均池化	93.3
全局最大值池化	92.3
局部平均池化	92.5
局部最大值池化	92.6
全局平均 + 局部平均	93.4
全局平均 + 局部最大值	93.6
全局最大值 + 局部平均	93.1
全局最大值 + 局部最大值	92.6

本文中提出的三大改进策略也可以应用于其他精细识别任务, 针对类间差小、种类繁多的目标, 设计出具有针对性的卷积神经网络模型。

References

- Su Ti, Yang Ming, Wang Chun-Xiang, Tang Wei, Wang Bing. Classification and regression tree based traffic merging for method self-driving vehicles. *Acta Automatica Sinica*, 2018, **44**(1): 35–43
(苏锦, 杨明, 王春香, 唐卫, 王冰. 一种基于分类回归树的无人车汇流决策方法. *自动化学报*, 2018, **44**(1): 35–43)
- Song D, Tharmarasa R, Kirubarajan T, Fernando X N. Multi-vehicle tracking with road maps and car-following models. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **19**(5): 1375–1386
- Yu Y, Wang J, Lu J T, Xie Y, Nie Z X. Vehicle logo recognition based on overlapping enhanced patterns of oriented edge magnitudes. *Computers and Electrical Engineering*, 2018, **71**: 273–283
- Hu C P, Bai X, Qi L, Wang X G, Xue G J, Mei L. Learning discriminative pattern for real-time car brand recognition. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(6): 3170–3181
- Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 2014, **115**(3): 211–252
- Lowe D G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004, **60**(2): 91–110
- Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *Proceedings of the 2005 Computer Vision and Pattern Recognition*. California, USA: IEEE, 2005. 886–893
- Luo Jian-Hao, Wu Jian-Xin. A survey on fine-grained image categorization using deep convolutional features. *Acta Automatica Sinica*, 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. *自动化学报*, 2017, **43**(8): 1306–1318)
- Liu W Y, Wen Y D, Yu Z D, Li M, Raj B, Song L. SpheroFace: Deep hypersphere embedding for face recognition. In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, HI, USA: IEEE, 2017.
- Mao J Y, Xiao T T, Jiang Y N, Cao Z M. What can help pedestrian detection? In: *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Hawaii, HI, USA: IEEE, 2017. 6034–6043
- Tang P, Wang X G, Huang Z L, Bai X, Liu W Y. Deep patch learning for weakly supervised object classification and discovery. *Pattern Recognition*, 2017, **71**: 446–459
- Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 2012, **25**(2): 1097–1105
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv Preprint, 2014, arXiv: 1409.1556
- He K M, Zhang X Y, Ren S Q, Sun J. Deep residual learning for image recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, Nevada, USA: IEEE, 2016. 770–778
- Dong Z, Wu Y W, Pei M T, Jia Y D. Vehicle type classification using a semisupervised convolutional neural network. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(4): 2247–2256
- Hsieh J W, Chen L C, Chen D Y. Symmetrical SURF and its applications to vehicle detection and vehicle make and model recognition. *IEEE Transactions on Intelligent Transportation Systems*, 2014, **15**(1): 6–20
- Liao L, Hu R M, Xiao J, Wang Q, Xiao J, Chen J. Exploiting effects of parts in fine-grained categorization of vehicles. In: *Proceedings of the 2015 IEEE International Conference on Image Processing*. Quebec City, Canada: IEEE, 2015. 745–749
- Biglari M, Soleimani A, Hassanpour H. Part-based recognition of vehicle make and model. *IET Image Processing*, 2017, **11**(7): 483–491
- He H S, Shao Z Z, Tan J D. Recognition of car makes and models from a single traffic-camera image. *IEEE Transactions on Intelligent Transportation Systems*, 2015, **16**(6): 3182–3192
- Lin Y L, Morariu V I, Hsu W, Davis L S. Jointly optimizing 3D model fitting and fine-grained classification. In: *Proceedings of the 2014 European Conference on Computer Vision*. Zurich, Switzerland: Springer, Cham, 2014. 466–480
- Krause J, Stark M, Deng J, Li F F. 3D object representations for fine-grained categorization. In: *Proceedings of the 2013 IEEE International Conference on Computer Vision Workshops*. Sydney, Australia, NSW: IEEE 2014. 554–561
- Sochor J, Herout A, Havel J. Boxcars: 3D boxes as CNN input for improved fine-grained vehicle recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, Nevada, USA: IEEE, 2016. 3006–3015
- Gao Y B, Lee H J. Local tiled deep networks for recognition of vehicle make and model. *Sensors*, 2016, **16**(2): 226

- 24 Le Q V, Ngiam J, Chen Z, Chia D J H, Pang W K, Ng A Y. Tiled convolutional neural networks. *Advances in Neural Information Processing Systems*, 2010: 1279–1287
- 25 Yu S Y, Wu Y, Li W, Song Z J, Zeng W H. A model for fine-grained vehicle classification based on deep learning. *Neurocomputing*, 2017, 257: 97–103
- 26 Yu Ye, Jin Qiang, Fu Yun-Xiang, Lu Qiang. Fine-grained classification of car models using Fg-CarNet convolutional neural network. *Acta Automatica Sinica*, 2018, 44(10): 1864–1875
(余焯, 金强, 傅云翔, 路强. 基于 Fg-CarNet 的车辆型号精细分类研究. 自动化学报, 2018, 44(10): 1864–1875)
- 27 Hu B, Lai J H, Guo C C. Location-aware fine-grained vehicle type recognition using multi-task deep networks. *Neurocomputing*, 2017, 243(Supplement C): 60–68
- 28 Fang J, Zhou Y, Yu Y, Du S D. Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture. *IEEE Transactions on Intelligent Transportation Systems*, 2017, 18(7): 1782–1792
- 29 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014, 15(1): 1929–1958
- 30 Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, et al. Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. NY, USA: ACM, 2014. 675–678
- 31 Wang J J, Yang J C, Yu K, Lv F J, Huang T S, Gong Y H. Locality-constrained linear coding for image classification. In: Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition, California, USA: IEEE, 2010. 3360–3367
- 32 Krause J, Gebu T, Deng J, Li L J, Li F F. Learning features and parts for fine-grained recognition. In: Proceedings of the 22nd International Conference on Pattern Recognition (ICPR). Stockholm, Sweden, 2014. 26–33
- 33 Liu X, Xia T, Wang J, Yang Y, Zhou F, Lin Y Q. Fine-grained recognition with automatic and efficient part attention. arXiv Preprint, 2016, arXiv, 1603. 06765
- 34 Wang Y M, Choi J, Morariu V I, Davis L S. Mining discriminative triplets of patches for fine-grained classification. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, Nevada, USA: IEEE, 2016. 1163–1172
- 35 Krause J, Jin H L, Yang J C, Li F F. Fine-grained recognition without part annotations. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 5546–5555
- 36 Yang L J, Luo P, Chen C L, Tang X O. A large-scale car dataset for fine-grained categorization and verification. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 3973–3981
- 37 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D et al. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 1–9
- 38 Huang G, Liu Z, Maaten L V D, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Hawaii, HI, USA: IEEE, 2017. 2261–2269



余焯 合肥工业大学计算机与信息学院副教授. 2010 年获合肥工业大学计算机科学与技术专业博士学位. 主要研究方向为图像处理, 计算机视觉, 虚拟现实与可视化. 本文通信作者.

E-mail: yuye@hfut.edu.cn

(**YU Ye** Associate professor at the School of Computer Science and Information, Hefei University of Technology. She received her Ph. D. degree from Hefei University of Technology in 2010. Her research interest covers image processing, computer vision, and virtual reality and visualization. Corresponding author of this paper.)



傅云翔 合肥工业大学计算机与信息学院硕士研究生. 2016 年获得合肥工业大学计算机与信息学院学士学位. 主要研究方向为图像处理, 计算机视觉与深度学习.

E-mail: yasinifu@mail.hfut.edu.cn

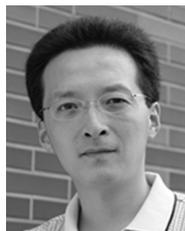
(**FU Yun-Xiang** Master student at the School of Computer Science and Information, Hefei University of Technology. He received his bachelor degree from Hefei University of Technology in 2016. His research interest covers image processing, computer vision, and deep learning.)



杨昌东 合肥工业大学计算机与信息学院硕士研究生. 2017 年获得黄山学院信息工程学院学士学位. 主要研究方向为图像处理, 计算机视觉与深度学习.

E-mail: 2017170766@mail.hfut.edu.cn

(**YANG Chang-Dong** Master student at the School of Computer Science and Information, Hefei University of Technology. He received his bachelor degree from Huangshan University in 2017. His research interest covers image processing, computer vision, and deep learning.)



路强 合肥工业大学计算机与信息学院副教授. 2010 年获合肥工业大学计算机科学与技术专业博士学位. 主要研究方向为可视化, 协同计算与图像处理.

E-mail: luqiang@hfut.edu.cn

(**LU Qiang** Associate professor at the School of Computer Science and Information, Hefei University of Technology. He received his Ph. D. degree from Hefei University of Technology in 2010. His research interest covers visualization, cooperative computing, and image processing.)