基于 Fg-CarNet 的车辆型号精细分类研究

余 烨 [1] 金 强 [1] 傅云翔 [1] 路 强 [1]

摘 要 车辆型号识别在智能交通系统、涉车刑侦案件侦破等方面具有十分重要的应用前景. 针对车辆型号种类繁多、部分型号区分度小等带来的车辆型号精细分类困难的问题, 采用车辆正脸图像为数据源, 提出一种多分支多维度特征融合的卷积神经网络模型 Fg-CarNet (Convolutional neural networks for car fine-grained classification, Fg-CarNet). 该模型根据车正脸图像特征分布特点, 将其分为上下两部分并行进行特征提取, 并对网络中间层产生的特征进行两个维度的融合, 以提取有区分度的特征, 提高特征表达能力, 通过使用小卷积核以及全局均值池化, 使在网络分类准确度提高的同时降低了网络模型参数大小. 在 CompCars 数据集上进行验证, 实验结果表明, Fg-CarNet 提取的车辆特征在保证网络模型参数最小的同时, 车辆型号识别率达到最高, 实现了最好的分类效果.

关键词 车辆型号精细分类, 卷积神经网络, 多维度特征融合, 分块并行

引用格式 余烨, 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

DOI 10.16383/j.aas.2017.c170109

Fine-grained Classification of Car Models Using Fg-CarNet Convolutional Neural Network

YU Ye[1] JIN Qiang[1] FU Yun-Xiang[1] LU Qiang[1]

Abstract Car model recognition has very important application in intelligent transportation systems and vehicle-related criminal case detection. A multi-branch and multi-dimension feature fusion convolutional neural network (CNN) model, Fg-CarNet (convolutional neural networks for car fine-grained classification), is proposed. This model uses car frontal face images as data source, and aims to solve the classification difficulty caused by the wide variety of car models and little differentiation between some models. Based on the image feature distribution characteristic of frontal face images, the Fg-CarNet divides them into upper parts and lower parts to extract features in parallel, and then merges the features generated by middle layers of the network to extract more distinguishing features. Through using small convolution kernel and global average pooling, the classification accuracy of Fg-CarNet is improved and at the same time the size of network parameters is reduced. With CompCars dataset, experiments are carried out. The results show that the proposed method can achieve the highest recognition accuracy while keeping the smallest size of network parameters, i.e, the method can achieve the best classification result.

Key words Fine-gained classification of car models, convolutional neural network (CNN), multi-dimension feature fusion, block parallel

Citation 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

车牌、车标、车型等车辆特征识别是智能交通领域的重要研究分支, 在违法犯罪车辆跟踪[1]、交通流量统计[2]、收费站自动收费等方面发挥着重要的作用. 由于盗牌、无牌、污损车牌车辆的存在, 使得车牌识别不能发挥应有的作用. 车标所占比例过小, 用于描述车标特征的像素有限, 且实际卡口监控系统中的车标定位是一个尚无成熟解决方案的难题, 此外, 车标也容易被替换、污损, 因此, 车标可以作为车辆识别的辅助特征, 而不是唯一特征. 目前, 车辆唯一不易被伪装的特征是 "车型", 其发挥的作用不可小觑. 车型识别为事故逃逸、套牌、假牌车辆等的发现和追踪提供辅助手段, 为交通管理执法部门提供重要判断依据, 因此具有巨大的研究价值和应用前景.

对车型的理解有两种, 车辆类型和车辆型号, 车

收稿日期 2017-02-28 录用日期 2017-08-02
Manuscript received February 28, 2017; accepted August 2, 2017
安徽省重点研究与开发计划项目 (1604d0802009), 安徽省自然科学基金 (1708085MF158), 安徽高校省级自然科学研究项目 (KJ2014ZD27) 资助
Supported by Provincial Key Research and Development Program of Anhui (1604d0802009), Natural Science Foundation of Anhui Province (1708085MF158), and Provincial Natural Science Research Projects in Anhui Universities (KJ2014ZD27)
本文责任编委 赖剑煌
Recommended by Associate Editor LAI Jian-Huang
1. 合肥工业大学计算机与信息学院 合肥 230009
1. School of Computer and Information, Hefei University of Technology, Hefei 230009

辆类型一般包含客车、卡车、轿车等分类, 车辆型号指车辆的具体款式, 例如, 轿车中的大众品牌, 里面有帕萨特、途观等型号. 车辆型号识别是一个典型的精细分类问题, 其研究面临很大的挑战, 这是因为: 1) 车辆型号种类繁多. 当前路面上常见的车辆型号达 1 000 多种以上, 种类越多意味着分类难度越大. 2) 部分车辆型号区分度小. 同一品牌中不同子型号的车辆存在外观差异度极小的情况, 且不同品牌中两种子型号的车辆也存在外观极其相似的情况. 差异度小意味着类间方差小, 需要提取更深层次、更抽象的特征才能实现其分类. 3) 图像受环境干扰大. 实际卡口监控系统中获取的车辆图像, 由于受周围环境、天气、光照等影响, 干扰较大, 增加了车辆型号识别的难度.

传统基于手工设计的特征提取方式往往由于关注点片面、抽象能力不足, 无法提取有区分度的特征对车辆型号进行描述. 随着 Hinton 等[3] 提出无监督逐层训练方法以来, 为训练深层神经网络提供了思路, 且随着近年来计算机硬件的发展, 运算能力大大增加, 使训练更深层次的神经网络成为可能, 从而掀起了一股深度学习的热潮. 卷积神经网络作为一种多层前馈深度学习模型, 由于其可以直接以图像为输入, 自动学习特征, 从而避免了手工设计特征抽象能力、区分度不足的问题, 在计算机视觉、图像处理等众多领域得到了广泛应用, 如目标跟踪[4]、图像分类[5]、语义分割[6] 和行为识别[7] 等.

在车辆型号识别中, 考虑到实际智能交通系统中获取的监控图像大部分是车辆正脸的照片, 且车辆正脸部分是车辆最具有区分度的区域, 因此, 本文以车辆正脸照片为数据源, 对车辆型号进行精细分类研究.

针对车辆正脸图像的特点进行分析, 由于车辆正脸图像特征分布不均, 尤其体现在上下两部分上, 为避免相同卷积核操作对上下两部分特征提取粒度不同, 以造成有用特征损失的问题, 针对车辆型号的精细分类, 设计了一种多分支多维度特征融合的卷积神经网络模型 Fg-CarNet (Convolutional neural networks for car fine-grained classification, Fg-CarNet). Fg-CarNet 具有如下特点: 1) 针对车辆正脸图像上下两部分设计不同的子网络, 并对上层子网络单独设置辅助损失函数, 使卷积神经网络能够在车脸图像不同区域提取不同的且具有区分度的特征. 2) 利用上下子网络中提取的不同特征的组合, 以及多尺度卷积核特征的组合, 进一步提高了卷积神经网络的识别准确率. 3) 网络中主要使用小尺寸卷积核来优化网络结构, 同时加入全局均值池化的方法, 使得网络在准确率提高的同时降低了网络参数的数量, 从而降低了网络过拟合的风险, 提高了网络的实用性.

1 相关工作

与车辆身份相关的识别工作主要分为三类: 车辆类型、车辆品牌和车辆型号 (如图 1 所示). 三者的分类精度由粗到细, 随着分类精细度的增加, 分类的难度越来越大, 实用性也越来越高. 车辆类型识别即根据车辆的大小、形状特征, 将其归为轿车、面包车、客车、卡车等类别, 主要用于高速路口自动收费、违规车辆检测等. 常用的识别方法有: 1) 基于视频中相邻几帧出现的车辆尺寸和线性特征, 结合车道宽度, 进行车辆类型的判断[8]. 2) 基于车辆模型的先验知识, 通过对各种环境下、各个角度、各种类型模型的匹配和参数调整来进行车辆类型的识别[9]. 3) 使用特征描述子如 GABOR[10]、Harris 角点[11] 和 SIFT[12] 等提取车辆特征 (这里的车辆特征不仅包含视觉特征, 也包含声音信号特征), 并使用分类器如 SVM[13] 进行分类识别. 4) 使用卷积神经网络的方法自动学习车辆特征, 并用于车辆类型的分类[14]. 由于不同类型车辆类间方差较大, 且类型种类较少, 现阶段对车辆类型的识别已取得了很好的效果, 识别率最高可达 96.1 %[14].

车辆品牌识别又称车辆制造商识别, 即判断车辆是大众、奥迪、起亚还是丰田. 由于车辆标志是车辆品牌的唯一特征, 因此目前车辆品牌主要基于车标的类型来进行判断. 文献 [15] 提出使用增强 SIFT 特征对车标进行识别, 在 1 200 张属于 10 种车辆品牌的车标样本库上进行测试, 平均识别率可达 91 %. 实际监控系统中获取的车标图像受光照影响较大, 为改善光照对车标识别的影响, 文献 [16] 提出了一种点对特征对车标进行识别, 在对 20 种车标进行识别时, 最高平均识别率可达 95.7 %. 然而, 文献 [16] 并未对车标的定位进行研究. 事实证明, 大部分车标识别方法都对车标定位有较高依赖, 定位好坏直接影响最后的识别结果. 为避免这一问题, 文献 [17] 提出使用多示例学习方法为每种品牌车辆找到最具有区分性的特征, 可以是车灯、车标、车辆边缘部位特征或其组合, 从而进行车辆品牌识别, 在包含 30 种车辆品牌数据集上进行测试, 识别率可以达到 94.66 %.

与车辆类型识别和品牌识别相比, 车辆型号识别难度更大, 这不仅因为车辆型号种类繁多, 更是因为不同车辆型号之间差异度过小, 即类间方差小, 难以找到有区分性的特征. 文献 [18] 提出了一种新的级联分类器集合方案, 通过加入拒绝策略, 尽量减少误分类样本带来的损失, 提高了车型分类的准确率. 文献 [19] 将车辆按照固定的网格进行划分, 并对每个网格提取 SURF 特征点和 HOG 特征

图 1 车辆身份相关的识别工作

Fig. 1 Recognition related to vehicle identity

后训练弱分类器, 最后用贝叶斯平均将这些弱分类器集成实现对车辆的分类, 在 29 类车型上取得了 99 % 的准确率. 对于这种精细分类问题, 部分方法通过寻找目标上具有区分度的细节部位来进行分类[20−22]. 文献 [23] 使用具有部位标定的车辆数据集训练 DPM 模型来定位车辆的关键区域, 再对每个区域提取特征后实现对车型的精细分类. 由于二维图像所包含的信息有限, 部分学者提出提取车辆的三维结构信息来辅助车辆的精细分类[24−25], 利用车辆三维模型提供的车辆视角、车体各部位位置信息等, 提高车辆精细分类的准确率. 随着深度卷积神经网络在图像分类中的成功应用[5], 人们开始尝试使用它来解决精细分类问题. 在这类方法中, 通常利用标定好的大量数据来训练一个深度卷积神经网络模型, 基于此模型从输入图像中提取高度抽象且区分度高的特征, 在网络的全连接层中将提取的特征连接成特征向量, 然后用分类器对得到的特征向量进行分类. 文献 [26] 在公布一个大型车辆数据集的基础上, 针对卡口监控系统中拍摄的车辆正脸图像, 使用 Alexnet[5]、Overfeat[27] 和 GoogLeNet[28] 等深度卷积神经网络 (Convolutional neural network, CNN) 模型对车辆型号的精细分类进行了研究. 文献 [29] 提出了一种多任务训练网络的方法, 将 Softmax Loss 和 Triplet Loss 共同作为训练目标, 在同一 CNN 网络中进行训练, 通过在损失函数中嵌入分级标签信息, 确保不同等级的类内方差小于类间方差, 从而提高了车辆型号精细分类的准确率. 文献 [30] 将车型的精细识别问题视为一个逐步求精的过程, 提出了一个由粗到精的卷积神经网络模型, 通过融合整体的特征以及局部具有区分度区域的特征实现车辆型号的精细分类.

在利用深度卷积神经网络解决图像分类问题时, 一个好的 CNN 模型提取的特征对提高分类效果起到了至关重要的作用. 为提高识别效果, 经典的卷积神经网络模型主要通过增加模型的深度和宽度来提取抽象程度更高的特征, 如 Alexnet、GoogLeNet 和 VGG-16[31] 等. 然而更深的网络意味着需要更多的计算资源和更多的样本来训练, 会造成训练难度

的增加, 从而导致网络性能下降的问题. 本文针对真实卡口场景下的车辆型号精细识别问题, 利用车辆正脸图像特征分布的特点, 设计了一个适用于车辆型号精细识别的模型 Fg-CarNet, 在使用较少网络权值的前提下, 获得了较好的车型识别效果.

2　车辆精细分类卷积神经网络模型

2.1　Fg-CarNet 模型结构

卷积神经网络是为识别二维形状而特殊设计的多层感知器[32]. 典型卷积神经网络的输入层为原始图像; 隐层由卷积层和池化层组合交替排列组成, 以减少网络的权值数量, 降低计算量; 为逐步建立网络空间和结构的不变性, 在卷积层后增加激活函数层以提高网络的非线性抽象能力; 将前面几层操作后获得的特征图在全连接层进行向量化, 并将提取的特征映射为标签, 根据标签进行物体类型的判断.

针对车辆正脸图像特征分布的特点, 在基本卷积神经网络结构的基础上进行改进, 考虑到车辆图像上下两部分特征的差异, 设计了一种适用于车辆型号精细分类的卷积神经网络模型 Fg-CarNet, 其模型结构如图 2 所示, 图中数字表示特征图的数量, 即卷积核的数量, N 表示网络最终输出的类别数.

Fg-CarNet 的输入图像为分割出的车辆正脸图像, 沿图像中线将其分割为上下两部分, 分别使用两段不同的分支网络 UpNet 和 DownNet 提取上下两部分的特征, 然后用特征融合网络 FusionNet 对 UpNet 和 DownNet 中提取的特征进行多维度融合, 进一步进行抽象并控制最终得到的特征规模, 最后利用全局均值池化代替传统的全连接层, 利用分类器得到网络的输出.

UpNet 是为了提取车辆图像上半部分粗轮廓特征而设计的一个浅层分支网络. 它由四个卷积层组成, 每个卷积层后都紧跟着一个 ReLU[5] 激活函数层、一个 Batch Normalize[33] 层和一个最大值池化层. ReLU 激活函数层进行特征映射; Batch Normalize 层对输出的结果进行规范化, 以加速网络的收敛; 最大池化层则实现对特征的降维. 本文将卷积层、激活函数层、Batch Normalize 层和最大值池化层四层连在一起的结构定义为一个网络的基本单元, 则 UpNet 由 4 个这样的基本单元组成.

DownNet 是针对车辆下半部分图像设计的深层子网络, 是 UpNet 结构的扩展, 由于车辆图像下半部分纹理特征密集, 包含更多有区分度的信息, 是车辆型号精细分类的关键特征所在, 因此, DownNet 在 UpNet 四个基本单元的卷积层后增加了一层卷积核大小为 1×1 的卷积层, 在深层次卷积层对浅层次卷积层学习到的特征进行整合前, 对浅层次的特征进行进一步的抽象, 提高了网络的表达能力.

FusionNet 首先将 UpNet 和 DownNet 第一个基本单元和最后一个基本单元提取的特征图进行上下组合, 得到两组完整的车辆特征图, 如图 2 中上部虚框线所示. 针对第一个基本单元的合并特征图, 使用一个基本单元进行特征提取, 对应 FusionNet 中的第二层. 此基本单元卷积核尺寸和步长与 UpNet 中使用的卷积核尺寸和步长不同, 详见表 1. 将 FusionNet 第二层得到的特征图和 UpNet、DownNet 中第四个基本单元组合得到的特征图叠加在一起, 如图 2 中下部虚框线所示, 再用一个基本单元对融合的特征图学习进一步特征提取, 并用两个 1×1 卷积层进行进一步的特征抽象和降维, 最后利用全局均值池化得到最后的分类特征.

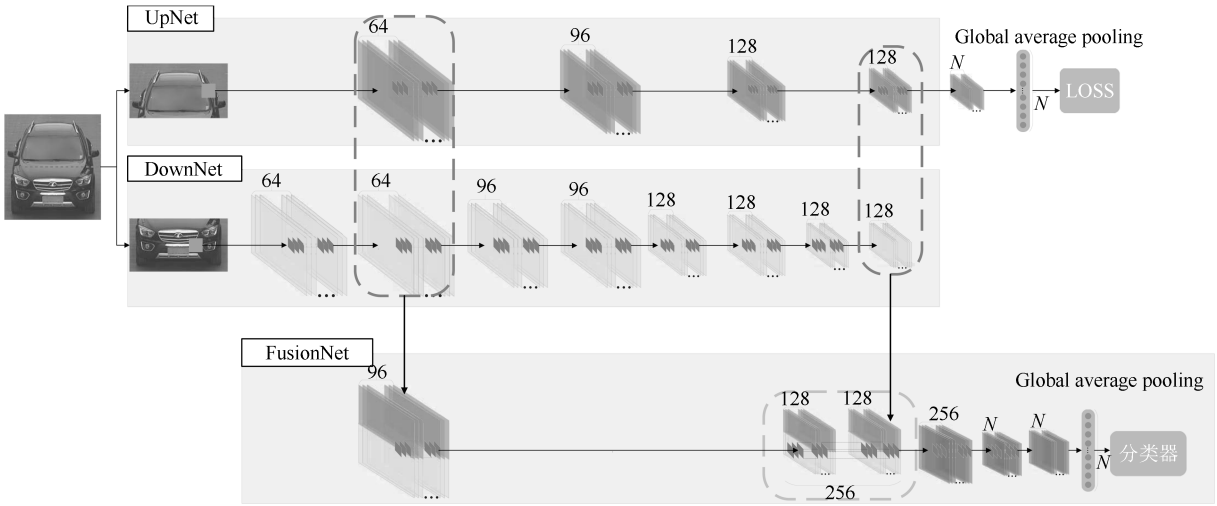


图 2　Fg-CarNet 网络结构示意图

Fig. 2　Network structure diagram of the Fg-CarNet

表 1 Fg-CarNet 模型结构参数

Table 1 Structural parameters of the Fg-CarNet

子网络	层编号	类型	卷积核尺寸/步长	池化类型	池化尺寸和步长	输出尺寸 (深度 × 长度 × 高度)
UpNet	1	Convolution/BN	$5\times5/2$	Max pooling	$3\times3/2$	$64\times64\times32$
	2	Convolution/BN	$3\times3/1$	Max pooling	$3\times3/2$	$96\times32\times16$
	3	Convolution/BN	$3\times3/1$	Max pooling	$3\times3/2$	$128\times16\times8$
	4	Convolution/BN	$3\times3/1$	Max pooling	$3\times3/2$	$128\times8\times4$
DownNet	1	Convolution	$5\times5/2$	–	–	$64\times128\times64$
	2	Convolution/BN	$1\times1/1$	Max pooling	$3\times3/2$	$64\times64\times32$
	3	Convolution	$3\times3/1$	–	–	$96\times64\times32$
	4	Convolution/BN	$1\times1/1$	Max pooling	$3\times3/2$	$96\times32\times16$
	5	Convolution	$3\times3/1$	–	–	$128\times32\times16$
	6	Convolution/BN	$1\times1/1$	Max pooling	$3\times3/2$	$128\times16\times8$
	7	Convolution	$3\times3/1$	–	–	$128\times16\times8$
	8	Convolution/BN	$1\times1/1$	Max pooling	$3\times3/2$	$128\times8\times4$
FusionNet	1	Concat	–	–	$96\times32\times32$	
	2	Convolution	$3\times3/2$	Max pooling	$2\times2/2$	$128\times8\times8$
	3	Concat	–	–	–	$128\times8\times8$
	4	Concat	–	–	–	$256\times8\times8$
	5	Convolution/BN	$3\times3/1$	Max pooling	$3\times3/2$	$256\times4\times4$
	6	Convolution/Drop	$1\times1/1$	–	–	$281\times4\times4$
	7	Convolution/Drop	$1\times1/1$	–	–	$281\times4\times4$
	8	Convolution	$1\times1/1$	Global pooling	–	$281\times1\times1$

Fg-CarNet 网络的具体参数设置如表 1 所示, Convolution 表示单独的卷积层, Convolution/BN 表示一个卷积层加一个 BatchNormalize 层, Convolution/Drop 表示一个卷积层加一个 Drop 层, 此外, 每个卷积层后都有一个 ReLU 激活函数 (表 1 中未明确列出). FusionNet 中第一层的 Concat 层, 输入为 UpNet 中第一层与 DownNet 中第二层输出特征图, 融合方式为对应层上下组合; FusionNet 中第三层的 Concat 层, 输入为 UpNet 中第四层和 DownNet 中第八层输出特征图, 融合方式为对应层上下组合; FusionNet 中第四层的 Concat 层, 其输入为 FusionNet 中第二层卷积层和第三层 Cancat 层输出特征图, 融合方式为特征层叠加.

2.2 分块特征提取

实际卡口监控系统中, 摄像头通常位于车辆上方, 斜向下对迎面而来的车辆进行拍摄. 被拍摄到的车辆部位从下到上依次包括车脸 (车大灯、车标、雾灯、散热器格栅和车牌)、引擎盖、挡风玻璃和部分车顶, 形成车辆正脸图像. 车脸是车辆特征最密集, 最具有区分度的部位, 具有丰富的纹理、形状特征, 通常位于正脸图像下方. 车辆正脸图像中除车脸外的其他部位也提供了丰富的轮廓、形状和位置信息, 这些特征可以作为车脸特征的一个补充.

将车辆正脸图像分为上下两部分, 则上下两部分的特征存在如下关系:

1) 下半部分的车脸所包含的特征多且细, 区分度高, 纹理特征密集, 车灯、隔热栅等形状特征明显. 而上半部分的图像以车辆挡风玻璃、车顶等为主, 主要体现为轮廓特征, 以及一些能反映细节的位置信息, 纹理特征不明显.

2) 在夜晚及一些特殊环境中, 车辆正脸上下两部分所处的光环境也存在较大差异, 且下半部分的车牌、散热器格栅、大灯及车标等通常使用特殊的材质, 对光的反射也与车辆上半部分差距较大, 这使得上下两部分在成像时存在亮度差异, 导致了上下两部分特征的区别.

如果使用卷积神经网络模型直接对整幅车辆正脸图像进行训练, 则在训练过程中, 卷积核会偏向于提取更有区分度的车脸部分特征来降低损失函数的值, 最终学习到的网络权重使网络中的神经元对车辆正脸图像的上下两部分激活不平衡, 导致上部分

图像的特征提取不足, 甚至丢弃, 整体学习到的特征区分度不足, 从而降低准确率.

图 3 是利用车辆正脸图像训练 AlexNet、GoogLeNet 和本文提出的 Fg-CarNet 网络模型后, 分别用白天和夜晚的两张图像作为输入, 将正向传播过程中卷积层的激活值进行可视化的结果. 为确保可视化结果具有对比性, 这里均选择三个网络结构中, 经过一次特征提取和映射且输出大小相近的层进行可视化, 分别为: Alexnet 的第一个卷积层、GoogLeNet 的第一个池化层和 Fg-CarNet 中两个子网络的第一个池化层. 分别提取上述层中特征图的前 16 张, 分为上下两部分进行可视化.

图 3 最左边一列为白天和夜晚的 2 张输入图像, 右边 3 列分别为不同网络模型提取的特征图可视化结果, 图像中灰度值越高, 越亮的部分表明神经元的激活程度越高. 从整体上看, 夜间车辆图像在神经网络中传播时, 神经元的激活度明显低于白天的车辆图像. 从单张图像在某个模型中提取的特征图来看, 车脸部分对应的神经元激活度明显高于车正脸上半部分对应的神经元, 这也证明了前文所述的观点: 车正脸具有上下两部分特征分布不均匀的特点. 从同一张图像在不同神经网络模型中提取的特征图可以看出, AlexNet 中神经元的激活度明显低于另外两种模型, 特别是车正脸上半部分对应的神经元, 大部分都处于激活度低或没有被激活的状态; GoogLeNet 中车脸部位对应神经元的激活度较 AlexNet 有明显提高, 但对于夜间车脸上半部分图像, 其神经元的激活度依然较低; 而 Fg-CarNet 由于是对上下两部分分开处理的, 所以车脸上半部分图像也能提取出有效的特征, 即使是在夜间, 也能保证神经元有较高的激活度. 针对此类特征分布具有明显空间结构且各部分特征粗细粒度不同的分类问题, 一个好的特征提取器需要能够统筹兼顾, 将各种有用的信息聚合起来构成最终特征. 鉴于此, 我们采用了分块特征提取的策略, 即针对车辆图像的上下两部分, 分别构建分支网络 UpNet 和 DownNet, 用于对上下两部分的特征分别进行提取. 为在特征丰富的车脸部分提取更具区分度的特征, DownNet 使用了比 UpNet 更长的网络. 在训练阶段, 对 UpNet 添加了额外的辅助 loss, 强制 UpNet 能学到更具区分度的特征, 使得网络能够从车脸上半部分图像中提取到足够丰富的特征. 如图 3 最右边一列图像所示, 与 AlexNet 和 GoogLeNet 相比, 本文提出的 Fg-CarNet 较好地改善了对车正脸上半部分特征提取的结果, 即使是在夜间, 车正脸上半部分对应的神经元依然有很好的激活度. 值得注意的是, 图 3 中, AlexNet 和 GoogLeNet 在同一车辆正面图像上下两部分进行特征可视化时, 使用的是同样的一组卷积核, 而本文的 Fg-CarNet 由于设计了两个子网络分别提取车辆图像的上下两部分特征, 所以各自用

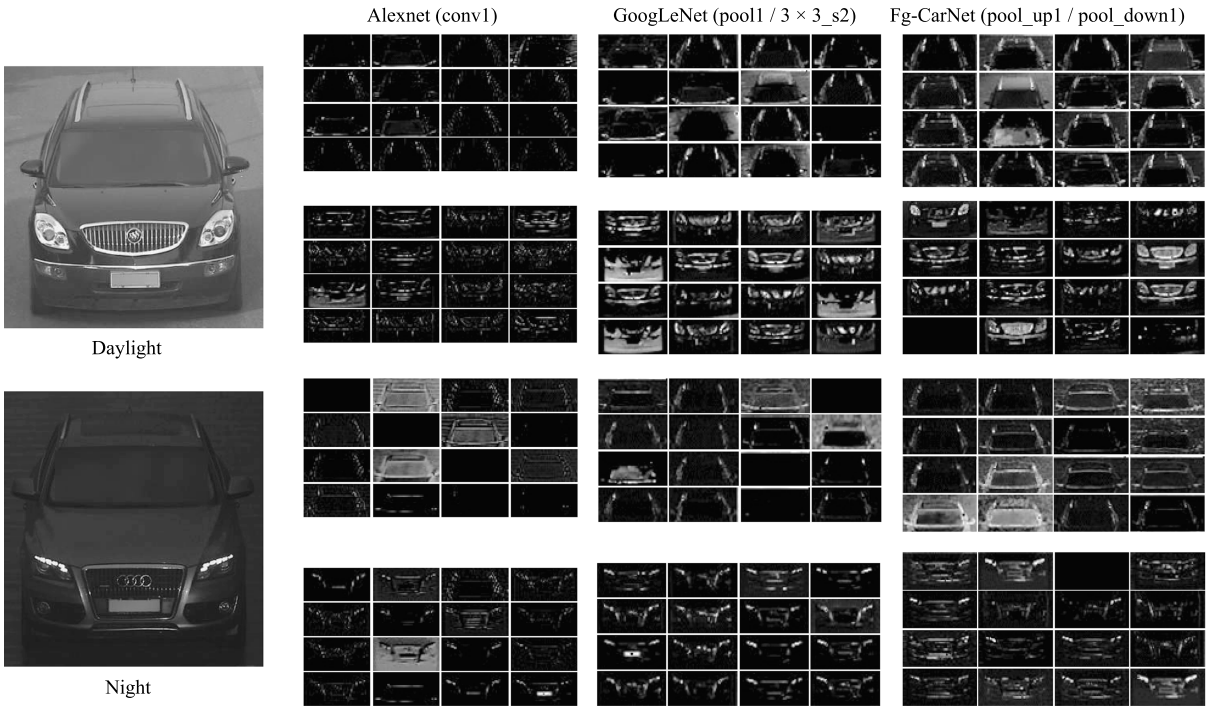


图 3　三类神经网络模型中层激活值可视化图

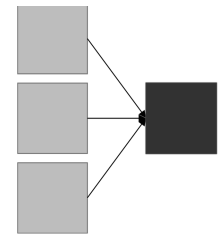
Fig. 3　Visualization of the layer activations in three neural network models

的卷积核是不同的.

2.3 多维度特征融合

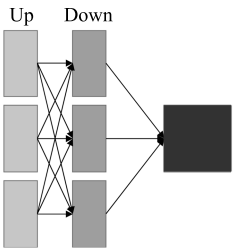
2.3.1 上下子特征融合

如前所述, 为对卡口车辆正脸图像上下两部分分别提取不同的特征, 训练阶段 Fg-CarNet 学习了两个分支网络对其进行特征提取, 之后会将两个分支提取的特征合并为一个整体作为车辆的特征. 设训练过程中某次正向传播 UpNet 得到的特征维度为 $N \times C \times H_{\text{up}} \times W$, DownNet 得到的特征维度为 $N \times C \times H_{\text{down}} \times W$, 其中 N 表示每次训练的 batch_size, C 表示通道数, 即特征图的个数, H_{up}, H_{down} 表示得到的特征图的高度, W 表示得到的特征矩阵的宽度, 则合并后的特征维度为 $N \times C \times (H_{\text{up}} + H_{\text{down}}) \times W$, 此处不仅是特征维度的增加, 由于上下两层采用了不同的卷积核提取特征, 因此, 此处更是一种特征的组合. 如图 4 所示, 传统卷积神经网络在正向传播过程中, 会将前一层卷积层产生结果的全部或部分作为输入, 而在 Fg-CarNet 中, UpNet 和 DownNet 提取的特征之间可以有多种组合方式, 设 UpNet 的特征图数量为 N_u, DownNet 的特征图数量为 N_d, 则可获得的组合数为 $N_u \times N_d$. 通过固定数量的卷积核得到多种组合的完整特征图, 提高了特征的利用率. 针对车辆型号精细识别, 这种组合方式可以将高激活度的车辆上半部分特征图与车辆下半部分特征图进行组合, 使得整个车辆特征图上的激活值都处于较高状态.



(a) 传统特征传播

(a) Traditional feature propagation



(b) 特征组合传播

(b) Feature combination propagation

图 4 层之间特征传播方式示意图

Fig. 4 Feature propagation between layers

2.3.2 多尺度卷积特征融合

传统 CNN 网络结构通常在每一层设置相同尺寸的卷积核, 对输入进行计算后得到输出并向下一层传递. 不同于这种结构, 本文对上下两个子网络提取的特征分别进行了低层 (靠近输入层的层) 和高层 (靠近输出层的层) 的融合, 即在第 2.3.1 节上下子特征融合的基础上, 针对低层融合后的特征, 使用一层具有较大卷积核的卷积层进行一次特征提取和降维, 并将得到的特征再次与高层融合后的特征进行叠加, 共同作为后面层的输入.

多尺度卷积特征融合 (如图 5 所示) 的优点可以从两方面来分析: 1) Fg-CarNet 网络结构使用不同尺寸的卷积核进行特征提取, 对同一输入, 一方面使用小尺寸的卷积核, 逐层进行特征提取和映射, 进行细粒度特征的提取; 另一方面, 使用大尺寸的卷积核, 直接进行粗粒度特征的提取, 保留更多的车辆轮廓信息. 粗粒度和细粒度特征的融合, 从不同尺度尽可能的保留了车辆正脸图像的特征, 提高了网络的特征表达能力. 2) 如图 5 中 Loss 标注线所示, 在训练网络的过程中, 训练误差可以从多个分支反向传播回低层卷积层, 优化了信息的流动, 可以有效避免因网络过深产生的梯度消散, 及导致低层卷积层得不到很好训练的问题.

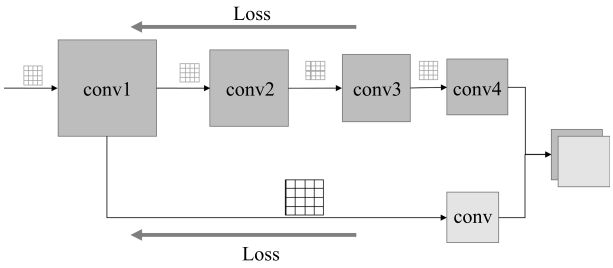


图 5 多尺度卷积特征融合

Fig. 5 Multiscale convolution feature fusion

3 实验结果与分析

3.1 实验数据集

为验证 Fg-CarNet 神经网络模型对卡口图像中车辆精细特征提取的有效性, 本文在文献 [26] 提出的 CompCars 数据集上对模型提取的特征进行了分类测试. CompCars 数据集是一个大规模的车辆数据集, 包含来自互联网和实际卡口监控系统 1 716 种型号的 208 826 张车辆图像. Fg-CarNet 神经网络模型主要针对卡口拍摄车辆正面图像进行车辆型号的精细特征提取, 故使用了 CompCars 中的卡口监控数据集进行测试, 这部分覆盖了夜晚、雨天和雾天等复杂环境下共 281 个型号的 44 481 张车辆图像,

其部分样例如图 6 所示. 本文使用其中的 70 % 作为训练集, 其他的 30 % 作为测试集.

图 6　CompCars 中监控数据集样例

Fig. 6　Sample images of the surveillance data in CompCars

3.2　实验环境及设置

实验的硬件环境如下: CPU 为 Intel Core i7-6 700 K; 内存为 32 GB; 显卡为 Nvidia GTX TITAN X; 显存为 12 GB. 实验所有模型均在开源框架 CAFFE[34] 下实现, CUDA 版本为 8.0.

在对 UpNet 和 DownNet 中基本单元提取特征进行融合的阶段, 要求两个网络产生的特征图相互匹配, 因此 Fg-CarNet 采用的输入尺寸为 256 × 256, 使其更易于控制网络传播过程中特征图的尺寸, 便于网络的设计. 而实验用到的其他网络皆按网络原有的要求输入为 224 × 224. 在训练和测试阶段, 各网络都对输入数据按照文献 [5] 中方法做了除尺寸外相同的预处理, 例如: 在训练 Fg-CarNet 网络时, 首先将数据集中所有样本的大小归一化为 290 × 290, 分别获得图像中心和四个拐角处共 5 张大小为 256 × 256 的图像, 再对获得的图像进行镜像操作, 获得其水平翻转后的图像, 由此, 基于每个样本可以获得 10 张训练图像. 最后, 所有图像均减去整个数据集的均值. 测试阶段仅通过裁剪获得图像中心大小为 256 的图像, 并减去图像均值. 实验所用网络模型的优化策略均为带有动量的分块随机梯度下降方法, 其中动量设置为 0.9, batch_size 设置为 128, 初始学习率设置为 0.001. 采用分步降低的策略, 每 100 k 次迭代降低 10 倍, 训练阶段共迭代 300 k 次, 故整个训练阶段学习率降低两次. 在测试阶段, 本文同样对测试样本进行扩增, 截取图像中心和四个拐角处共 5 张大小为 256 × 256 的图像, 再对获得的图像进行镜像操作, 获得其水平翻转后的图像, 对扩增后的 10 个样本分别求置信度后取平均作为最后的结果.

3.3　Fg-CarNet 在 CompCars 上的性能评估

表 2 显示了使用本文提出的 Fg-CarNet 深度卷积神经网络模型以及经典的 AlexNet, GoogLeNet 和 Network in network[35] (NIN) 深度神经网络模型在 CompCars 数据集上提取车辆特征, 并使用朴素贝叶斯分类器、KNN 分类器、逻辑回归分类器、随机森林分类器、SVM 分类器和 Softmax 分类器对车辆进行分类的结果. 从表 2 可以看出, AlexNet 和 NIN 网络模型提取的特征在车辆精细识别方面的识别率较低, GoogLeNet 提取特征的表现则要高

表 2　卷积神经网络模型在 CompCars 上使用不同分类器的识别率

Table 2　Recognition rate of different CNN models using different classifiers on CompCars

分类器	各模型识别率			
	AlexNet (%)	GoogLeNet (%)	NIN (%)	Fg-CarNet (%)
朴素贝叶斯	91.10	96.95	86.06	98.42
KNN	93.41	98.35	92.78	98.78
逻辑回归	96.08	98.39	95.91	98.76
随机森林	82.96	95.61	74.60	93.52
SVM	96.02	98.33	96.23	98.78
Softmax	97.73	98.50	96.51	98.89

很多, 而本文提出的 Fg-CarNet 模型提取的特征, 除了使用随机森林分类器外, 使用其他分类器的准确率都要高于 GoogLeNet, 达到了 98 % 以上. 总体上看, Fg-CarNet 的识别准确率是最高的. 从分类器效果上来分析, Softmax 分类器在各网络模型提取特征上的分类效果都是最好的.

卷积神经网络中参数的数量反映了模型的拟合能力, 参数越多越容易过拟合, 泛化能力也越差, 此外, 参数越多需要的内存也越多, 网络的适用性会降低. 使用 CAFFE 框架训练卷积神经网络后生成一个保存网络结构参数的文件, 表 3 以 CAFFE 生成的模型参数文件大小来反映各个网络的参数规模, 从表 3 可以看出, Fg-CarNet 模型的参数数量远低于其他三种模型, 与 AlexNet 相比, 参数数量下降了近 40 倍; 与性能表现相近的 GoogLeNet 相比, 参数数量减少了近 6 倍; 与同样采用了全局均值池化的 Network in Network 模型相比, 参数规模也降低了 1 倍, 然而性能却得到了大大提升.

表 3 各神经网络模型参数的大小

Table 3 The size of each CNN model parameters

神经网络模型	模型参数大小 (MB)
AlexNet	232.1
GoogLeNet	44.7
NIN	12.8
Fg-CarNet	6.3

3.4 与其他车型分类算法的比较

针对真实卡口拍摄车辆正脸图像的精细分类, 与其他针对多视角的车型分类方法有所不同, 分类性能也有所差异, 因此本文仅与针对卡口图像中车辆正脸图像进行精细分类的相关工作进行比较, 比较结果如表 4 所示. 由于 CompCars 监控数据集中, 属于各类别车型的图像数量差距较大, 最少的仅 14 张, 而最多的可达 565 张, 为避免会忽略样本数量少的类别中识别率不佳的情况, 采用文献 [30] 建议的用两种评估方式分别对实验结果进行评估. 各自的计算公式如下:

$$Accuracy1 = \frac{\sum_{i=1}^{N} t_i}{N} \tag{1}$$

$$Accuracy2 = \frac{\sum_{i=1}^{N} \frac{t_i}{n_i}}{N} \tag{2}$$

其中, t_i 为每类中正确预测的样本的数量, n_i 为每类样本的数量, N 为样本的类别数.

表 4 相关工作的识别结果

Table 4 Report results of some related works

序号	模型方法	类别数	准确率 1 (%)	准确率 2 (%)
1	NIN	281	96.51	95.25
2	AlexNet	281	97.73	96.72
3	GoogleNet	281	98.50	97.90
4	Zhang[18]	281	–	83.78
5	Hsieh 等[19]	281	–	51.70
6	Fang 等[30]	281	98.63	98.29
7	Ours	281	98.89	98.27

表 4 中, 第 1 ∼ 3 行是采用经典卷积神经网络对 CompCars 监控数据集进行分类的结果, 从中可以看出, GoogLeNet 的识别结果最好, 其准确率 1 达到了 98.5 %, 准确率 2 达到了 97.9 %. 第 4 ∼ 6 行是与文献 [18−19, 30] 中车型分类算法在 CompCars 监控数据集上分类性能的比较, 其中, 文献 [18−19] 的实验结果均来自文献 [30]. 为了实验的公平性, 文献 [30] 在 CompCars 监控数据集上复现了文献 [18−19] 的实验, 实验结果表明, 其所提方法在大规模车辆型号精细分类问题上性能不佳. 文献 [30] 提出的方法在 CompCars 监控数据集上取得了较高的准确率, 准确率 1 达到了 98.63 %, 准确率 2 达到了 98.29 %. 本文提出的方法与文献 [30] 方法相比, 准确率 1 更高, 准确率 2 基本持平, 说明本文方法在准确率方面优于文献 [30] 方法. 此外, 本文提出的 Fg-CarNet 模型是一个端到端的模型, 可以直接快速地实现车辆型号的精细分类, 而且由于使用了全局均值池化, 大大降低了网络中参数的数量, 提高了网络的可使用性.

3.5 分块融合的性能评估

第 2.2 节指出, 对车辆正脸图像分成上下两部分进行特征提取, 可以强制特征不明显的上半部分区域也提取出有助于车型分类的特征, 并与下半部分区域提取的特征进行融合, 以提高车辆型号精细分类的准确率. 为验证将车辆正脸图像分成两部分进行特征提取对车型识别带来的影响, 本文设计了一组对比实验, 实验结果如表 5 所示. Fg-CarNet-Up 和 Fg-CarNet-Down 分别为以卡口车辆正脸图像的上半部分和下半部分作为输入的网络. 为了保证公平性, 减少因为网络深度带来的影响, Fg-CarNet-Up 和 Fg-CarNet-Down 分别由 Fg-CarNet 网络中的 FusionNet 删除融合部分, 保留基本的特征提取部分后与 UpNet 和 DownNet 相连接后获得. 从实验结果可以看出, 仅使用车辆的上半部分正脸图

像, 准确率 1 为 93.37 %, 准确率 2 为 89.78 %. 而仅使用特征更丰富的车辆下半部分正脸图像, 准确率 1 达到了 97.38 %, 但其准确率 2 较准确率 1 下降较多, 这是因为当用 Fg-CarNet-Down 进行分类时, 存在样本较少的类别其准确率较低, 从而导致了准确率 2 的降低. 整体上, Fg-CarNet-Up 的准确率均低于 Fg-CarNet-Down 的准确率, 这也证明了本文的观点, 即车辆上半部分正脸图像具有一定的区分性, 但特征不及下半部分正脸图像明显. Fg-CarNet-Whole 与 Fg-CarNet-Down 模型结构相同, 但 Fg-CarNet-Whole 是以整张车辆正脸图像作为输入, 准确率 1 的结果达到了 98.02 %, 准确率 2 的结果达到了 97.84 %, 与单独使用上半部分或下半部分图像进行车型分类相比, 准确率得到了明显提高. 而 Fg-CarNet 由于对上下两部分单独采用不同的子网络进行特征提取, 并将各自提取的特征进行多维度融合, 增强了网络对车辆的特征描述能力, 最终准确率 1 和准确率 2 与上述几种相比均有提高.

表 5　分块融合的性能比较

Table 5　Performance comparison of block fusion

模型	准确率 1 (%)	准确率 2 (%)
Fg-CarNet-Up	93.37	89.78
Fg-CarNet-Down	97.38	93.82
Fg-CarNet-Whole	98.02	97.84
Fg-CarNet	98.89	98.27

3.6　网络特征的可视化分析

为进一步分析分块特征提取的效果, 本文将 GoogLeNet, AlexNet 和 Fg-CarNet 基于 CompCars 测试集提取的特征, 使用 t-SNE[36−37] 方法降维到二维进行可视化, 可视化结果如图 7 所示, 图中一个点代表一个测试样本, 同样灰度点表示同一类样本. 由于通常卷积神经网络的最后一层是用于将特征映射到特定类别, 因此此处我们选择最后一层的前一层所提取特征进行可视化. 图 7 (a) 中, 样本点整体呈现一种聚类趋势, 但重叠度较高, 如图中左下角部分, 基本混杂在一起, 这说明 UpNet 针对车辆上半部分图像学习到了可用于车型分类的特征, 但区分程度不够; 图 7 (b) 中, 各类样本能够较好地聚在一起, 具有明显的区分界限, 但类间距离不够大, 这也证明了第 2.2 节所述, 车辆正脸下半部分图像包含更多有区分度的特征, 更有利于车型分类; 从图 7 (c) 中对 AlexNet 提取特征进行可视化的结果可以看出, 虽然 AlexNet 提取的特征整体类间差距较大, 但类内差距也很大, 这并不有利于分类; 图 7 (d) 所示是对 GoogLeNet 提取特征进行可视化的结果, 可以看出各类样本点被很好地区分开来, 且同一类样本点紧凑的聚集在一起, 然而类间差距依然不够大; 而融合了 UpNet 和 DownNet 的 Fg-CarNet 提取的特征如图 7 (e) 所示, 各类区分明显, 类间差较大, 说明 Fg-CarNet 提取的车辆特征能够较好地将类与类之间区分开来, 同时类内样本聚合度较高, 能够实现较好的分类效果.

3.7　不同特征融合性能评估

在多尺度卷积特征融合阶段, 不同层特征的组合可能会产生不同的分类结果, 为对不同组合下的分类性能进行评估, 分别对不同组合情况下的模型识别率进行测试, 测试结果如表 6 所示. 其中, 单元编号对应 UpNet 和 DownNet 中四个基本单元的编号. 从表 6 可以看出, 不同组合下模型的识别率区别不大, 第一和第四个基本单元特征图融合后的性能达到最优, 为 98.906 %. 模型 7 融合的参数最多, 但分类性能并不是最优, 融合了 3 个参数的模型 4, 5, 6, 其性能也没有模型 1 的性能高, 这说明并非融合的特征越多, 分类性能就越高. 融合的特征越多, 可能会导致特征冗余, 且网络中的模型参数也会随之增多.

4　结论

针对卡口图像中车辆型号精细分类问题进行研究, 提出了 Fg-CarNet 深度卷积神经网络模型, 以

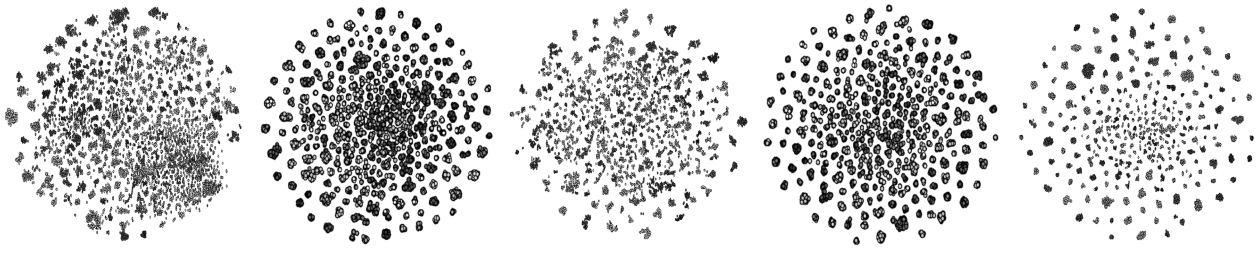


图 7　特征降维后可视化结果

Fig. 7　Visualization of features after dimension reduction

表 6 不同基本单元特征组合下的识别结果
Table 6 Recognition result based on different basic unit combinations

模型编号	单元编号				
	1	2	3	4	准确率 1
1	✓			✓	0.98906
2		✓		✓	0.98789
3			✓	✓	0.98843
4	✓	✓		✓	0.98835
5	✓		✓	✓	0.98901
6		✓	✓	✓	0.98882
7	✓	✓	✓	✓	0.98835

提取有区分度的特征, 提高车辆型号精细分类的准确性. Fg-CarNet 的主要特点是: 1) 采用分块并行的方式, 分别用 UpNet 和 DownNet 两个分支网络对车正脸图像的上下两部分进行特征提取, 提高特征提取的有效性; 2) 对提取的车正脸图像的上下两部分特征进行了两个维度的融合, 提高了特征的表达能力; 3) 网络使用小卷积核及全局均值池化代替了传统的全连接网络实现特征向结果的映射, 大大地降低了模型的参数规模. 实验结果表明, 本文提出的 Fg-CarNet 能够以较少的参数提取具有区分度的车辆精细特征, 在分类性能上表现优异, 具有实用价值. 此外, 本文提出的分区域特征提取和多维度特征融合的方法, 对其他不同区域间关联度低物体的精细分类问题也提供了思路.

References

1 Coifman B, Beymer D, McLauchlan P, Malik J. A real-time computer vision system for vehicle tracking and traffic surveillance. *Transportation Research, Part C: Emerging Technologies*, 1998, **6**(4): 271−288

2 Wu Cong, Li Bo, Dong Rong, Chen Qi-Mei. Detecting traffic parameters based on vehicle clustering from video. *Acta Automatica Sinica*, 2011, **37**(5): 569−576
(吴聪, 李勃, 董蓉, 陈启美. 基于车型聚类的交通流参数视频检测. 自动化学报, 2011, **37**(5): 569−576)

3 Hinton G E, Salakhutdinov R R. Reducing the dimensionality of data with neural networks. *Science*, 2006, **313**(5786): 504−507

4 Guan Hao, Xue Xiang-Yang, An Zhi-Yong. Advances on application of deep learning for video object tracking. *Acta Automatica Sinica*, 2016, **42**(6): 834−847
(管皓, 薛向阳, 安志勇. 深度学习在视频目标跟踪中的应用进展与展望. 自动化学报, 2016, **42**(6): 834−847)

5 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, Nevada, USA: ACM, 2012. 1097−1105

6 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015. 3431−3440

7 Zhu Yu, Zhao Jiang-Kun, Wang Yi-Ning, Zheng Bing-Bing. A review of human action recognition based on deep learning. *Acta Automatica Sinica*, 2016, **42**(6): 848−857
(朱煜, 赵江坤, 王逸宁, 郑兵兵. 基于深度学习的人体行为识别算法综述. 自动化学报, 2016, **42**(6): 848−857)

8 Hsieh J W, Yu S H, Chen Y S, Hu W F. Automatic traffic surveillance system for vehicle tracking and classification. *IEEE Transactions on Intelligent Transportation Systems*, 2006, **7**(2): 175−187

9 Zhang Z X, Tan T N, Huang K Q, Wang Y H. Three-dimensional deformable-model-based localization and recognition of road vehicles. *IEEE Transactions on Image Processing*, 2012, **21**(1): 1−13

10 Ji P J, Jin L W, Li X T. Vision-based vehicle type classification using partial Gabor filter bank. In: Proceedings of the 2007 IEEE International Conference on Automation and Logistics. Jinan, China: IEEE, 2007. 1037−1040

11 Li J, Zhao W Z, Guo H. Vehicle type recognition based on Harris corner detector. In: Proceedings of the 2nd International Conference on Transportation Engineering. Chengdu, China: Southwest Jiaotong University, 2009. 3320−3325

12 Kang Wei-Xin, Cao Yu-Ting, Sheng Zhuo, Li Peng, Jiang Peng. Harris corner and SIFT feature of vehicle and type recognition. *Journal of Harbin University of Science and Technology*, 2012, **17**(3): 69−73
(康维新, 曹宇亭, 盛卓, 李鹏, 姜澎. 车辆的 Harris 与 SIFT 特征及车型识别. 哈尔滨理工大学学报, 2012, **17**(3): 69−73)

13 Qi X X, Ji J W, Han X W, Yuan Z H. An approach of passive vehicle type recognition by acoustic signal based on SVM. In: Proceedings of the 3rd International Conference on Genetic and Evolutionary Computing. Guilin, China: IEEE, 2009. 545−548

14 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

15 Psyllos A P, Anagnostopoulos C N E, Kayafas E. Vehicle logo recognition using a SIFT-based enhanced matching scheme. *IEEE Transactions on Intelligent Transportation Systems*, 2010, **11**(2): 322−328

16 Yu Ye, Nie Zhen-Xing, Jin Qiang, Wang Jiang-Ming. Vehicle logo recognition based on randomly sampled pixel-pair feature from foreground-background skeleton areas. *Journal of Image and Graphics*, 2016, **21**(10): 1348−1356
(余烨, 聂振兴, 金强, 王江明. 前背景骨架区域随机点对策略驱动下的车标识别方法. 中国图象图形学报, 2016, **21**(10): 1348−1356)

17 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

18 Zhang B L. Reliable classification of vehicle types based on cascade classifier ensembles. *IEEE Transactions on Intelligent Transportation Systems*, 2013, **14**(1): 322−332

19 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

20 Pandey G, McBride J R, Eustice R M. Ford campus vision and lidar data set. *The International Journal of Robotics Research*, 2011, **30**(13): 1543−1552

21 Xiao T J, Xu Y C, Yang K Y, Zhang J X, Peng Y X, Zhang Z. The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 842−850

22 Göring C, Rodner E, Freytag A, Denzler J. Nonparametric part transfer for fine-grained recognition. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Columbus, OH, USA: IEEE, 2014. 2489−2496

23 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 (ICIP). Quebec City, QC, Canada: IEEE, 2015. 745−749

24 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, 2014. 466−480

25 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 (ICCVW). Sydney, NSW, Australia: IEEE, 2014. 554−561

26 Yang L J, Luo P, Change Loy C, 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 (CVPR). Boston, USA: IEEE, 2015. 3973−3981

27 Sermanet P, Eigen D, Zhang X, Mathieu M, Fergus R, LeCun Y. Overfeat: integrated recognition, localization and detection using convolutional networks. arXiv: 1312.6229, 2014.

28 Szegedy C, Liu W, Jia Y Q, Sermanet P, Reed S, Anguelov D, Rabinovich A. Going deeper with convolutions. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA: IEEE, 2015. 1−9

29 Zhang X F, Zhou F, Lin Y Q, Zhang S T. Embedding label structures for fine-grained feature representation. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, Nevada, USA: IEEE, 2016. 1114−1123

30 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

31 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2015.

32 Haykin S O. *Neural Networks and Learning Machines* (3rd edition). Upper Saddle River, NJ, USA: Pearson, 2009.

33 Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning. Lille, France: PMLR, 2015, **37**: 448−456

34 Jia Y Q, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T. Caffe: convolutional architecture for fast feature embedding. In: Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, Florida, USA: ACM, 2014. 675−678

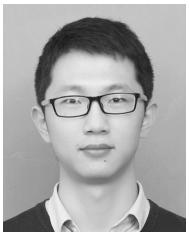
35 Lin M, Chen Q, Yan S C. Network in network. arXiv: 1312.4400, 2014.

36 Van Der Maaten L, Hinton G. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 2008, **9**(11): 2579−2605

37 Van Der Maaten L. Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research*, 2014, **15**(1): 3221−3245

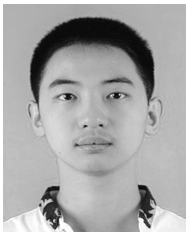
余 烨 合肥工业大学计算机与信息学院副教授. 2010 年获得合肥工业大学博士学位. 主要研究方向为图像处理, 计算机视觉, 虚拟现实与可视化. 本文通信作者. E-mail: yuye@hfut.edu.cn

(**YU Ye** Associate professor at the School of Computer 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, virtual reality and visualization. Corresponding author of this paper.)



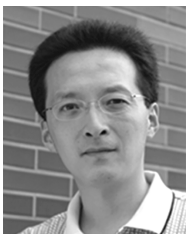
金 强 合肥工业大学计算机与信息学院硕士研究生. 2015 年获得合肥工业大学学士学位. 主要研究方向为图像处理, 计算机视觉与模式识别.
E-mail: ksstrong@mail.hfut.edu.cn

(**JIN Qiang** 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 2015. His research interest covers image processing, computer vision, and pattern recognition.)



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: yasinfu@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.)



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 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.)