



(12)发明专利

(10)授权公告号 CN 106250812 B

(45)授权公告日 2019.08.20

(21)申请号 201610563184.1

G06N 3/08(2006.01)

(22)申请日 2016.07.15

(56)对比文件

(65)同一申请的已公布的文献号
申请公布号 CN 106250812 A

CN 104657748 A,2015.05.27,

CN 105184271 A,2015.12.23,

US 2016148079 A1,2016.05.26,

CN 105404858 A,2016.03.16,

(43)申请公布日 2016.12.21

王茜等.基于深度神经网络的汽车车型识别.《图形图像》.2015,

(73)专利权人 汤一平

地址 310014 浙江省杭州市下城区新市街
64号朝晖六区71-3-501

邓柳等.基于深度卷积神经网络的车型识别研究.《计算机应用研究》.2016,第33卷(第3期),

(72)发明人 汤一平

审查员 刘欢

(74)专利代理机构 杭州天勤知识产权代理有限公司 33224

代理人 胡红娟

(51)Int.Cl.

G06K 9/00(2006.01)

G06K 9/62(2006.01)

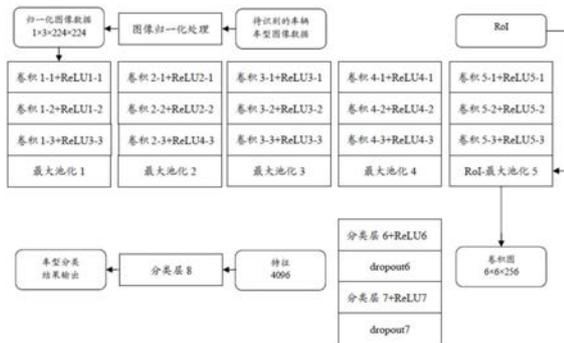
权利要求书4页 说明书18页 附图7页

(54)发明名称

一种基于快速R-CNN深度神经网络的车型识别方法

(57)摘要

本发明公开一种基于快速R-CNN深度神经网络的车型识别方法,主要包括无监督的深度学习、多层的CNN卷积神经网络、区域建议网络、网络共享、softmax分类器;实现了一个真正意义上的使用一个快速R-CNN网络实现端到端的车辆检测和识别的框架,并具有适用于车辆目标的形态多样性,光照变化多样性,背景多样性等环境下快速、高精度和鲁棒性的车辆子类识别。



1. 一种基于快速R-CNN深度神经网络的车型识别方法,其特征在于:包括一个用于深度学习和训练识别的VGG网络、一个用于提取出感兴趣区域的区域建议网络和一个用于车型分类的Softmax分类器;

所述的VGG网络,包括8个卷基层,3个全连接层,共计11层;8个卷基层包括5个组的卷积层、2个分类层提取图像特征和1个分类层分类特征;3个全连接层分别连接分类层6、分类层7和分类层8;

所述的区域建议网络,包括1个分类层、1个窗口回归层、1个计算分类损失的模块和1个计算窗口回归损失的模块,输出p个感兴趣的建议框;

所述的Softmax分类器,将提取到的输入数据特征与学习训练得到特征库数据进行比对,计算出每一个分类结果的概率,然后取概率最高的一个结果进行输出;

快速R-CNN深度神经网络,在所述的VGG网络的第5层末尾接入了所述的区域建议网络,使得所述的区域建议网络共享所述的VGG网络的前5层的底层特征提取过程与结果;

所述的VGG网络的第6层和第7层根据所述的区域建议网络输出的p个感兴趣的建议框内的图像特征进行卷积和ReLU处理,得到p个含有4096向量的特征图,接着分别送给分类层和窗口回归层进行处理,实现车辆图像的分割;所述的Softmax分类器对p个含有4096向量的特征图进行分类识别,得到车辆车型的分类结果。

2. 如权利要求1所述的车型识别方法,其特征在于:所述的区域建议网络,为了生成区域建议框,在所述的VGG网络的第5层末尾接入了所述的区域建议网络,即在第5层的卷积层输出的卷积特征映射图上滑动一个小网络,这个网络全连接到输入卷积特征映射的 $n \times n$ 的空间窗口上;每个滑动窗口映射到一个低维向量上,低维向量为256-d,每个特征映射的一个滑动窗口对应一个数值;这个向量输出给两个同级的全连接的层,一个窗口回归层和一个分类层;窗口回归层输出每一个位置上,9种推荐区域对应窗口平移缩放不变形,窗口回归层从256维特征中输出4个平移缩放参数,有4k个输出,即k个建议框的坐标编码;分类层从256维特征中输出属于前景和背景的概率,输出2k个建议框得分,即对每个建议框是车辆目标或非车辆目标的估计概率。

3. 如权利要求1或2所述的车型识别方法,其特征在于:区域建议网络的训练,给每个候选区域分配一个二进制的标签,即是否是车辆对象;这里分配正标签给两类候选区域:

(i) 与某个地面实况GT的包围区域有最高的交集并集之比IoU的重叠的候选区域;

(ii) 与任意地面实况GT包围区域的交集并集之比IoU大于0.7的重叠的候选区域;同时分配负标签给与所有GT包围区域的IoU比率都低于0.3的候选区域;

删去非正非负的候选区域;具体算法如下:

STEP31: 顺序读取训练集中的每张图;

STEP32: 对每个标定的真值候选区域,与其重叠比例最大的候选区域记为前景样本;

STEP33: 对STEP32剩余的候选区域,如果其与某个标定重叠,且IoU比例大于0.7,记为前景样本;如果其与任意一个标定的重叠比例都小于0.3,记为背景样本;

STEP34: 对STEP32和STEP33剩余的候选区域弃去不用;

STEP35: 对跨越图像边界的候选区域弃去不用。

4. 如权利要求3所述的车型识别方法,其特征在于:为了自动进行候选区域的筛选和区域位置精修,这里采用最小化目标函数;对一个图像的代价函数用公式(14)表示,

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) \quad (14)$$

式中, i 是一个批处理中候选区域的索引, N_{cls} 为分类层的归一化系数, N_{reg} 为窗口回归层的归一化系数, λ 为平衡权重, p_i 为车辆目标的预测概率, p_i^* 为 GT 标签, 如果候选区域为正 $p_i^* = 1$, 如果候选区域为负 $p_i^* = 0$; t_i 为一个向量, 表示预测的包围框的 4 个参数化坐标, t_i^* 为与正候选区域对应的 GT 包围框的坐标向量, L_{cls} 为分类的对数代价, L_{reg} 为回归对数代价, $L(\{p_i\}, \{t_i\})$ 为总的对数代价;

分类的对数代价 L_{cls} 由公式 (15) 计算,

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (15)$$

窗口回归对数代价 L_{reg} 由公式 (16) 计算,

$$L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) = R(\mathbf{t}_i - \mathbf{t}_i^*) \quad (16)$$

式中, R 为定义的鲁棒的代价函数, 属于 Smooth L1 误差, 对离群值不敏感, 用公式 (17) 计算,

$$R(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases} \quad (17)$$

公式 (14) 中的 $p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*)$ 这一项意味着只有正候选区域, 即 $p_i^* = 1$ 时才有回归代价, 其他情况由于 $p_i^* = 0$ 没有回归代价; 分类层和窗口回归层的输出分别由 $\{p_i\}$ 和 $\{t_i\}$ 组成, 这两项分别由 N_{cls} 和 N_{reg} 以及一个平衡权重 λ 归一化, 这里选择 $\lambda = 10$, $N_{cls} = 256$, $N_{reg} = 2400$, 通过这样的选择分类层和窗口回归层项是等权重的;

关于位置精修, 这里采用 4 个值, 中心坐标、宽度和高度, 计算方法如下,

$$\begin{aligned} t_x &= (x - x_a) / w_a, t_y = (y - y_a) / h_a, t_w = \log(w / w_a), t_h = \log(h / h_a), \\ t_x^* &= (x^* - x_a) / w_a, t_y^* = (y^* - y_a) / h_a, t_w^* = \log(w^* / w_a), t_h^* = \log(h^* / h_a) \end{aligned} \quad (18)$$

式中, x, y, w, h 分别表示包围框中心坐标、宽度和高度, x_a, y_a, w_a, h_a 分别表示候选区域中心坐标、宽度和高度, x^*, y^*, w^*, h^* 分别表示预测的包围框中心坐标、宽度和高度, t_x, t_y, t_w, t_h 分别表示经位置精修后的包围框中心坐标、宽度和高度; $t_x^*, t_y^*, t_w^*, t_h^*$ 分别表示经位置精修后的预测的包围框中心坐标、宽度和高度; 用公式 (18) 的计算结果进行位置精修区域建议网络。

5. 如权利要求 1 所述的车型识别方法, 其特征在于: 所述的 VGG 网络, 在标签车辆图像数据上建立多层神经网络的方法, 分为两步, 一是每次训练一层网络, 二是调优, 使原始表示 X 向上生成的高级表示 r 和该高级表示 r 向下生成的 X' 一致;

卷积神经网络的前向传播过程, 上一层的输出即为当前层的输入, 并通过激活函数逐层传递, 因此整个网络的实际计算输出用公式 (4) 表示,

$$O_p = F_n(\cdots (F_2(F_1(XW_1)W_2)\cdots)W_n) \quad (4)$$

式中, X 表示原始输入, F_n 表示第 1 层的激活函数, W_n 表示第 1 层的映射权值矩阵, O_p 表示

整个网络的实际计算输出；

当前层的输出用公式 (5) 表示，

$$X^l = f^l (W^l X^{l-1} + b^l) \quad (5)$$

式中， l 代表网络层数， X^l 表示当前层的输出， X^{l-1} 表示上一层的输出，即当前层的输入， W^l 代表已经训练好的、当前网络层的映射权值矩阵， b^l 为当前网络的加性偏置， f^l 是当前网络层的激活函数；采用的激活函数 f^l 为纠正线性单元，即ReLU，用公式 (6) 表示，

$$f^l = \max((W^l)^T X^l, 0) = \begin{cases} (W^l)^T X^l & (W^l)^T X^l > 0 \\ 0 & (W^l)^T X^l \leq 0 \end{cases} \quad (6)$$

式中， l 代表网络层数， W^l 代表已经训练好的、当前网络层的映射权值矩阵， f^l 是当前网络层的激活函数；其作用是如果卷积计算结果小于0，则让其为0；否则保持其值不变。

6. 如权利要求5所述的车型识别方法，其特征在于：所述的VGG网络，前5层是一个深度卷积神经网络，该神经网络训练是一个反向传播过程，通过误差函数反向传播，利用随机梯度下降法对卷积参数和偏置进行优化调整，直到网络收敛或者达到最大迭代次数停止；

反向传播需要通过对带有标签的训练样本进行比较，采用平方误差代价函数，对于 c 个类别， N 个训练样本的多类别进行识别，网络最终输出误差函数用公式 (7) 来计算误差，

$$E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \quad (7)$$

式中， E^N 为平方误差代价函数， t_k^n 为第 n 个样本对应标签的第 k 维， y_k^n 为第 n 个样本对应网络预测的第 k 个输出；

对误差函数进行反向传播时，如公式 (8) 所示，

$$\begin{aligned} \delta^l &= (W^{l+1})^T \delta^{l+1} \times f'(u^l) \\ u^l &= W^l x^{l-1} + b^l \end{aligned} \quad (8)$$

式中， δ^l 代表当前层的误差函数， δ^{l+1} 代表上一层的误差函数， W^{l+1} 为上一层映射矩阵， f' 表示激活函数的反函数，即上采样， u^l 表示未通过激活函数的上一层的输出， x^{l-1} 表示下一层的输入， W^l 为本层映射权值矩阵；

通过误差反向传播后，得到各个网络层的误差函数 δ^l ，然后采用随机梯度下降法对本层映射权值矩阵 W^l 进行修改，再进行下一次迭代，直至网络达到收敛条件；在进行误差传递时需要先通过公式 (8) 中的上采样使得前后两层尺寸相同，再进行误差传递；

算法思想是：1) 首先逐层构建单层神经元，这样每次都是训练一个单层网络；2) 当所有层训练完后，使用wake-sleep算法进行调优；

深度学习训练过程具体如下：

STEP21: 使用自下而上的非监督学习，即从底层开始，一层一层的往顶层训练，学习车辆图像特征：先用无标签车辆图像数据训练第一层，训练时先学习第一层的参数，由于模型容量的限制以及稀疏性约束，使得得到的模型能够学习到数据本身的结构，从而得到比输入更具有表示能力的特征；在学习得到第 $l-1$ 层后，将 $l-1$ 层的输出作为第 l 层的输入，训练第 l 层，由此分别得到各层的参数；

STEP22: 自顶向下的监督学习，即通过带标签的车辆图像数据去训练，误差自顶向下传输，对网络进行微调。

7. 如权利要求1所述的车型识别方法,其特征在于:所述的VGG网络的前5层的模型初始化:分为数据准备、计算图像均值、网络的定义、训练和恢复数据5个步骤;

1) 数据准备;收集各类车型的图像数据,包括带有标签的车辆图像数据和通过卡口摄像机获取的车辆图像数据,并将带有标签的车辆图像数据作为训练图像数据;

2) 计算图像均值;从每张图片减去均值;

3) 网络的定义;定义xml标签路径、图片的路径、存放train.txt、val.txt、test.txt和trainval.txt文件的路径;

4) 训练;运行训练模块;

5) 恢复数据;删除前ReLU5的层,并将roi_pool5的bottom改为data和rois;

经过上述处理完成了车辆车型预训练的模型初始化工作。

8. 如权利要求7所述的车型识别方法,其特征在于:所述的区域建议网络利用了所述的VGG网络的前5层底层特征提取结果,即两个网络共享了所述的VGG网络的前5层底层特征,需要通过交替优化来学习优化共享的特征;具体算法如下:

STEP41:优化训练区域建议网络,用所述的数据准备、计算图像均值、网络的定义、训练和恢复数据5个步骤完成模型初始化,并用端到端的方式微调区域建议任务;

STEP42:用STEP41的区域建议网络生成的建议框,由快速R-CNN训练一个单独的检测网络,这个检测网络同样是由车辆车型预训练的模型初始化的,这时候两个网络还没有共享卷积层;

STEP43:用检测网络初始化区域建议网络训练,固定共享的卷积层,并且只微调区域建议网络独有的层,这时两个网络共享卷积层了;

STEP44:保持共享的卷积层固定,微调快速R-CNN的分类层;这样,两个网络共享相同的卷积层,最终构成一个统一的网络。

9. 如权利要求1所述的车型识别方法,其特征在于:车辆车型视觉识别流程如下;

STEP51:读取待识别图像;

STEP52:对待识别图像进行归一化处理,得到RGB三个不同颜色归一化的图像数据;

STEP53:将RGB三个不同颜色归一化的图像数据输入到三个CNN通道,经5层卷积处理得到车辆车型特征图像数据;

STEP54:对车辆车型特征图像数据通过区域建议网络生成的建议框,选取一个最高得分的建议框,即得到一个感兴趣区域,RoI;将这个RoI通过最大池化5层进行处理得到一个RoI的卷积图;

STEP55:将RoI的卷积图输出给两个同级的全连接的层处理后得到4096维的特征向量,作为softmax分类器的输入数据;

STEP56:对4096维的特征向量用softmax的分类回归分析得到车辆车型识别结果,识别出被测图像中的车辆属于哪一款车型。

一种基于快速R-CNN深度神经网络的车型识别方法

技术领域

[0001] 本发明涉及计算机技术、模式识别、人工智能、应用数学以及生物视觉技术在智能交通领域的应用,尤其涉及一种基于快速R-CNN深度神经网络的车型识别方法。

背景技术

[0002] 智能交通系统中的核心功能是对过往车辆的准确检测和正确的车型识别。当前对车辆检测分类技术的研究主要有两个重要技术:即车辆自动识别和车辆自动分类。

[0003] 车辆自动识别是利用车载设备与地面基站设备互识进行,该技术主要用于收费系统中,在一些技术发达国家使用范围较广,如美国的AE-PASS系统、日本的ETC系统,全球卫星GPS定位等。

[0004] 车辆自动分类是通过检测车辆本身固有的参数,在一定车辆分类标准下运用适当的分类识别算法,主动地对车辆进行分型,这一类技术应用比较广泛,已经有很多成熟的系统应用在实际生活中,该类技术可以通过射频微波、红光、激光、声表面波等方式来自动识别车辆信息,也可以使用视频图像处理的方式来识别车牌、车型等车辆信息。

[0005] 车辆自动分类比较成熟技术有环形线圈检测、激光红外线检测、超声波/微波检测、地磁检测等,但这几种方法各有优劣,优点是识别精确比较高,但缺点也很明显,主要缺点有施工和安装过程十分复杂,影响正常交通秩序,维护困难,主要设备易损坏,花费较大等。

[0006] 近年来,视频检测技术已成为智能交通领域最重要的信息采集手段,综合评比,将视频检测技术应用于高速公路和城市道路具有极大的实际应用价值,基于视频车型识别系统,将全面提升城市道路的信息采集和安全管理水平,在智能交通系统中会发挥越来越重要的作用。

[0007] 车型的视觉识别,国内外很多学者进行了相关研究。论文“Robert T.Collins, Alan J.Lipton,Hironobu Fujiyoshi,and Takeo Kanade.Algorithms For cooperative multisensor surveillance.In Proceedings of the IEEE”披露了一个路上移动目标的检测、跟踪、识别系统,用训练过的神经网络来识别运动目标是人、人群、车辆还是干扰,网络的输入特性量有目标的分散性度量、目标大小目标表面大小与摄影机监视区域大小的相对值。车辆又进一步区分为不同类型和颜色。论文“TieniuN.Tan and Keith D.Baker Efficient image gradient based vehicle localization.IEEE Transaction on Image Processing,2000,9(8):1343-1356.”描述了一种车辆定位和识别的方法,在一个小窗口内,该方法依据图像梯度进行。利用地面约束以及大部分车辆外形受两条直线约束的事实,可得到车辆的姿态。论文“George Fung,NelsonYung,and Granthajm Pang.Vehicle shape approximation from motion for visual traffic surveillance.In Proc.IEEE Conf.Intelligent Transport System,2001,608-613.”用高精度摄像机观察车辆的运动来估计车辆形状,通过估计特征点得到车辆轮廓。基本思想是高特征点的移动速度大于低特征点的移动速度,因为高特征点离摄像机近,车辆轮廓可

用与车辆识别。论文“Ferryman,A.Worrall,G.Sullivan,K.Baker,A generic deformable model for Vehicle recognition,Proceeding of British Machine Vision Conference,1995,127-136.”提出了一个参数化的可变形三维模板,该模板通过演变,据称是可适用于各种车辆识别。论文“Larry Davis,V.Philomin and R.Duralswami.Tracking humans from a moving platform.In Proc.Intematl on Conference on Patten Recognition,2000.”用变形模板来研究车辆识别,首先,建立目标车辆车头部分的侧视图以及正视图的变形模板。通过直方图交集,车辆的RGB直方图也必须比较,合适的车型模板边的点集也通过点集间的Hausdorff距离与其他车辆模板进行比较。上述技术基本上还需要人工来完成特征抽取,更大的问题是:1)受具体应用环境的影响太大,各类检测算法要求的条件太苛刻;2)车辆种类繁多但差别不大,没有明显的区别特征;3)受视觉变化的影响大,从不同角度所摄的汽车特征差别大;4)受自然环境影响太大,特别是光照影响,严重的光照反射使得车辆轮廓线不分明,颜色偏离、变化太大,难以辨认;5)汽车的外形更新太快,特征变化太快,使得算法适应性较差。国内在车型识别技术方面大多数仍停留在研究状态,如中国科学院、西安公路所、上海交通大学、西安交通大学、四川大学等一些研究成果。其关键问题是人类本身对车型识别过程的知识限制。

[0008] 用来进行车辆分类的特征需要对尺度、旋转以及一定的角度变换、光照变化都有很好的鲁棒性,前深度学习时代的计算机视觉技术一般都是人工花费大量时间和精力去设计合适的特征。为了能让计算机自动选择合适特征,人工神经网络就应运而生了,早在1999年就有国外的研究者利用神经网络对物体进行分类,包括了模糊神经网络和BP网络等方法;但由于其性能存在诸多问题,如难以解决模式识别中存在的复杂性和推广性的矛盾,神经网络虽然有强大的建模能力,但对诸如车型大规模图像进行分类,其巨大的参数空间使寻找优良的优化初始值较为困难等,因此被人们冷落了很长时间,直到深度学习的提出才又重新成为研究热点。

[0009] 在图像分类领域,对数量庞大的图像分类主要有两种方法:一种为提取每张照片的局部特征,将提取的特征进行聚类 and 编码得到一个高维向量,然后将其用分类器分类。其中编码的方法有视觉词袋模型编码,稀疏编码和费舍尔向量编码等,从目前研究结果来看费舍尔向量编码的性能较其他几种编码方式要好。另一种应用很广的图像分类方法是深度神经网络,深度学习是神经网络研究中的一个新的热点,其目的在于通过非监督的预训练为神经网络提供优良的参数初始值,通过贪婪的方式,一层一层的训练,对大规模图像分类取得非常好的效果。

[0010] 深度学习的概念开始引起人们的注意是在大约2006年前后,论文“Hinton,G.E.and R.R.Salakhutdinov,Reducing the dimensionality of data with neural networks.Science,2006.313(5786):504-507”提出了一种多层的前向神经网络能够逐层做高效的前期训练,采用无监督的限制性玻尔兹曼机对每一层进行训练,最后在利用有监督的后向传播做微调,为解决模式识别中存在的复杂性和推广性的矛盾提供了一种新的方法和思路,这样就拉开了深度学习时代的计算机视觉技术序幕。

[0011] 卷积神经网络,即CNN,是深度学习算法的一种,是专门处理图像领域中的模式识别,同时也是目前图像模式识别中成果最惊人的算法。卷积神经网络算法的好处在于训练模型的时候不需要使用任何人工特征,算法可以自动探索图像所隐含的特征。

[0012] 申请号为201610019285.2的中国专利申请公开了一种车型识别方法及系统,包括了机器训练生成分类器过程和对待测图片的判别过程,在生成分类器的过程中,基于车牌确定训练集图片中所需要的图像范围,将已经确定的图像范围划分区域,选定各区域内的特征信息,分别将选定的各自区域内所有特征信息投入机器训练生成一一对应各区域的分类器,通过已经生成的分类器对待测图片进行单区域判别,依据单区域判别结果再经过多区域置信度融合判断得到车型识别结果。该发明采用随机深林分类器进行车型识别,最大的问题是缺少无监督的学习过程。

[0013] 申请号为201510639752.7的中国专利申请公开了一种车型识别方法,所述方法包括:获取待检测图片;采用第一预设分类器对所述待检测图片进行检测;若所述待检测图片中含目标车辆,提取所述待检测图片中的目标车辆;对所述目标车辆进行对齐处理,以使得所述目标车辆的车头方向与所述目标区域的垂直方向之间的夹角小于预设阈值;对所述对齐处理之后的目标车辆进行特征提取,以得到M个特征,所述M为大于1的整数;采用第二预设分类器对所述M个特征进行分类;根据所述分类的结果确定所述目标车辆的车型。从某种意义上说,该发明仍然属于浅神经网络,在架构上就难以实现无监督的学习过程。

[0014] 申请号为201510071919.4的中国专利申请提出了一种基于卷积神经网络的车型识别方法,基于特征提取模块与车型识别模块,包括以下步骤:通过设计卷积与池化层,全连接层,分类器来构建车型识别的神经网络,其中卷积与池化层和全连接层用来提取车型特征,分类器用来车型分类识别;利用包含不同车型特征的数据库训练该神经网络,训练方式为带标签的数据进行的有监督的学习,并用随机梯度下降法进行权重参数矩阵和偏移量的调整;得到训练好的各层中的权重参数矩阵和偏移量,把它们对应地赋值给该神经网络中的各个层,则该网络具有车型特征提取及识别的功能。该发明缺少比较详细的实现细节,只是从概念上提出了基于卷积神经网络的车型识别方法,尤其是训练方式为带标签的数据进行的有监督的学习。车辆的外形图像量属于海量数据,要对这些图像数据进行标注是一个极其困难的事情;此外车辆的外形更新太快造成特征变化快,所以造成该发明中的一些算法难以满足实际应用。此外,实际路面采集到的车型图像复杂;包括了背景复杂,车辆之间的遮挡等干扰;若无对实际获得的图像进行分割处理,将会严重影响最终的识别结果。

[0015] 申请号为201510738852.5和201510738843.6的中国专利申请提出了一种基于深度费舍尔网络的车型识别方法,首先构建费舍尔网络的第0层,对有K种车型图像的数据库,提取出每种车型车辆图像的SIFT特征;然后构建费舍尔网络的第1层,对提取到的SIFT特征进行费舍尔向量编码,将编码后的向量在空间堆叠,再进行L2归一化和PCA降维;将第1层得到的特征表示进行费舍尔向量编码,通过符号平方根和L2归一化处理,形成费舍尔网络的第2层;最后将不同车型图像得到的全局特征表示用于线性支持向量机训练,得到具有K种车型类别的识别系统;对待识别车辆,使其通过费舍尔网络得到测试特征向量,导入识别系统即可识别出待识别车辆车型。该发明也有两个不足,其一是缺少深度学习过程,其二是缺少对图像进行车辆分割处理步骤。

[0016] 申请号为201510738540.4的中国专利申请提出了一种基于局部特征聚合描述符的车型识别方法,首先提取车型数据库中车辆图像的SIFT特征;然后对所有车辆图像的SIFT特征进行Kmeans聚类,形成K个聚类中心,得到具有K个视觉单词的字典;接着针对每张车型图像将每个SIFT特征赋给离其最近的视觉单词;统计每个视觉单词周围的SIFT特征向

量与当前视觉单词的残差累积量,得到当前车辆图片的局部特征聚合描述符;最后,将训练模块的n张车型图像的局部特征聚合描述符,通过量化编码,得到一个可索引的n类车型类别的编码图像库;而对测试车辆图像,同样提取其局部特征聚合描述符,作为查询向量,导入图像库进行索引,通过近似最近邻搜索方法进行匹配,识别出测试车辆车型。同样该发明也有两个缺陷,其一是缺少深度学习,其二是缺少对图像进行车辆分割处理步骤。

[0017] 卷积神经网络在大货车、小货车、大巴、面包车、SUV和轿车识别上还是比较成功的,但是在子类分类,如识别车的不同车型中的精度就远比不上大类分类的精度。通常来说子类图像识别的困难一般在于两点:

[0018] (1) 子类图像标注数据的获取十分困难,通常需要相关领域的专家进行标注。

[0019] (2) 子类图像识别存在大的类内差别,如在车型识别中,车不同的观察角度,以及小的类间差别。

[0020] 综上所述,采用卷积神经网络等深度神经网络技术对车型识别,目前尚存在着如下若干个棘手的问题:1) 如何从复杂的背景中准确分割出被测车辆的整体图像;2) 如何尽可能采用极少的标签图像数据来准确获得车辆车型的特征数据;3) 如何在车辆车型大类识别基础上还能识别出是哪一款、哪种颜色、哪个年代出的车;4) 如何通过深度学习来自动获取车辆车型的特征;5) 如何兼顾好识别精度和检测效率,同时尽可能减少训练和学习时间;6) 如何在分类器设计时,该分类器即能满足车辆车型子类的分类要求,又能在汽车的外形更新后无需再重新对整个网络进行训练学习;7) 如何设计一个真正意义上的使用一个CNN网络实现端到端的车辆检测和识别的框架;8) 如何降低天气条件的影响,增加系统的自适应性。

发明内容

[0021] 为了克服已有的车辆车型视觉识别技术中的自动化和智能化水平低、缺乏深度学习、难以对适应环境天气变化、难以准确提取出用于识别的车辆整体图像、难以用视觉方式对车辆车型子类进行识别分类、难以兼顾识别精度与时间和检测效率等不足,本发明提供一种基于快速R-CNN深度神经网络的车型识别方法,能有效提高车型视觉识别的自动化和智能化水平、能较好适应环境天气变化具有广泛的自适应性、能保证在较好的检测识别精度基础上有实时检测识别能力、能大大减少对标签车辆数据的依赖具有自动学习和提取车型特征的能力、能较好解决车型识别的复杂性和推广性的矛盾具有较佳的普适性。

[0022] 要实现上述发明内容,必须要解决几个核心问题:(1) 设计一种车辆对象的快速视觉分割算法;(2) 研发一种深度学习方法,实现无监督的车型特征提取;(3) 设计一种适用于上千种的类型子类的分类器,并具有扩展性;(4) 设计一个真正意义上的使用一个快速R-CNN网络实现端到端的车辆检测和识别的框架。

[0023] 本发明解决其技术问题所采用的技术方案是:

[0024] 一种基于快速R-CNN深度神经网络的车型识别方法,包括一个用于深度学习和训练识别的VGG网络、一个用于提取出感兴趣区域的区域建议网络和一个用于车型分类的Softmax分类器;

[0025] 所述的VGG网络,包括8个卷基层,3个全连接层,共计11层;8个卷基层中有5个组的卷积层、2个分类层提取图像特征、1个分类层分类特征;3个全连接层分别连接分类层6、分

类层7和分类层8;

[0026] 所述的区域建议网络,包括1个分类层、1个窗口回归层、1个计算分类损失的模块和1个计算窗口回归损失的模块,输出p个感兴趣的建议框;

[0027] 所述的Softmax分类器,将提取到的输入数据特征与学习训练得到特征库数据进行比对,计算出每一个分类结果的概率,然后取概率最高的一个结果进行输出;

[0028] 快速R-CNN深度神经网络,在所述的VGG网络的第5层末尾接入了所述的区域建议网络,使得所述的区域建议网络共享所述的VGG网络的前5层的底层特征提取过程与结果;

[0029] 所述的VGG网络的第6和第7层根据所述的区域建议网络输出的p个感兴趣的建议框内的图像特征进行卷积和ReLU处理,得到p个含有4096向量的特征图,接着分别送给分类层和窗口回归层进行处理,实现车辆图像的分割;另一方面,所述的Softmax分类器对p个含有4096向量的特征图进行分类识别,得到车辆车型的分类结果。

[0030] 所述的Softmax分类器,在学习训练期间将快速R-CNN中的学习结果作为softmax分类器的输入数据;Softmax回归是面向多类分类问题的Logistic回归,是Logistic回归的一般形式,适用于类别之间互斥的情况;假设对于训练集 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$,有 $y^{(1)} \in \{1, 2, \dots, k\}$,对于给定的样本输入x,输出一个k维的向量来表示每一种分类结果出现的概率为 $p(y=i|x)$,假设函数 $h(x)$ 如下:

$$[0031] \quad h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1|x^{(i)}, \theta) \\ p(y^{(i)} = 2|x^{(i)}, \theta) \\ \vdots \\ p(y^{(i)} = k|x^{(i)}, \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (11)$$

[0032] $\theta_1, \theta_2, \dots, \theta_k$ 是模型的参数,并且所有的概率和为1;加入规则项后的代价函数为:

$$[0033] \quad J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{l=1}^k \sum_{j=0}^n \theta_{lj}^2 \quad (12)$$

[0034] 代价函数对第j个类别的第1个参数的偏导数为:

$$[0035] \quad \nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[x^{(i)} (1\{y^{(i)} = j\} - p(y^{(i)} = j|x^{(i)}; \theta)) \right] + \lambda \theta_j \quad (13)$$

[0036] 最后,通过最小化 $J(\theta)$,实现softmax的分类回归,将分类回归结果保存到特征库中;

[0037] 在识别分类时,将提取到的输入数据特征与学习训练得到特征库数据进行比对,计算出每一个分类结果的概率,然后取概率最高的一个结果进行输出。

[0038] 所述的区域建议网络,为了生成区域建议框,在所述的VGG网络的第5层末尾接入了所述的区域建议网络,即在第5层的卷积层输出的卷积特征映射图上滑动一个小网络,这个网络全连接到输入卷积特征映射的 $n \times n$ 的空间窗口上;每个滑动窗口映射到一个低维向量上,低维向量为256-d,每个特征映射的一个滑动窗口对应一个数值;这个向量输出给两个同级的全连接的层;一个窗口回归层和一个分类层;窗口回归层输出每一个位置上,9种推荐区域对应窗口需要具有平移缩放不变性,窗口回归层从256维特征中输出4个平移缩放参数,有4k个输出,即k个建议框的坐标编码;分类层从256维特征中输出属于前景和背景的

概率,输出2k个建议框得分,即对每个建议框是车辆目标/非车辆目标的估计概率。

[0039] 区域建议网络的训练,给每个候选区域分配一个二进制的标签,即是否是车辆对象;这里分配正标签给两类候选区域:(i)与某个GT的包围区域有最高的交集并集之比,IoU,重叠的候选区域;(ii)与任意GT包围区域大于0.7的IoU交叠的候选区域;同时分配负标签给与所有GT包围区域的IoU比率都低于0.3的候选区域;删去非正非负的候选区域;具体算法如下:

[0040] STEP31:顺序读取训练集中的每张图;

[0041] STEP32:对每个标定的真值候选区域,与其重叠比例最大的候选区域记为前景样本;

[0042] STEP33:对STEP32)剩余的候选区域,如果其与某个标定重叠,IoU比例大于0.7,记为前景样本;如果其与任意一个标定的重叠比例都小于0.3,记为背景样本;

[0043] STEP34:对STEP32和STEP33剩余的候选区域弃去不用;

[0044] STEP35:对跨越图像边界的候选区域弃去不用。

[0045] 为了自动进行候选区域的筛选和区域位置精修,这里采用最小化目标函数;对一个图像的代价函数用公式(14)表示,

$$[0046] \quad L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) \quad (14)$$

[0047] 式中,i是一个批处理中候选区域的索引, N_{cls} 为分类层的归一化系数, N_{reg} 为窗口回归层的归一化系数, λ 为平衡权重, p_i 为车辆目标的预测概率, p_i^* 为GT标签,如果候选区域为正 $p_i^* = 1$,如果候选区域为负 $p_i^* = 0$; t_i 为一个向量,表示预测的包围框的4个参数化坐标, t_i^* 为与正候选区域对应的GT包围框的坐标向量, L_{cls} 为分类的对数代价, L_{reg} 为回归对数代价, $L(\{p_i\}, \{t_i\})$ 为总的对数代价;

[0048] 分类的对数代价 L_{cls} 由公式(15)计算,

$$[0049] \quad L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (15)$$

[0050] 窗口回归对数代价 L_{reg} 由公式(16)计算,

$$[0051] \quad L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) = R(\mathbf{t}_i - \mathbf{t}_i^*) \quad (16)$$

[0052] 式中,R为定义的鲁棒的代价函数,属于Smooth L1误差,对离群值不敏感,用公式(17)计算,

$$[0053] \quad R(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases} \quad (17)$$

[0054] 公式(14)中的 $p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*)$ 这一项意味着只有正候选区域,即 $p_i^* = 1$ 时才有回归代价,其他情况由于 $p_i^* = 0$ 没有回归代价;分类层和窗口回归层的输出分别由 $\{p_i\}$ 和 $\{t_i\}$ 组成,这两项分别由 N_{cls} 和 N_{reg} 以及一个平衡权重 λ 归一化,这里选择 $\lambda = 10$, $N_{cls} = 256$, $N_{reg} = 2400$,通过这样的选择分类层和窗口回归层项差不多是等权重的;

[0055] 关于位置精修,这里采用4个值,中心坐标、宽度和高度,计算方法如下,

$$\begin{aligned}
 [0056] \quad & t_x = (x - x_a) / w_a, t_y = (y - y_a) / h_a, t_w = \log(w / w_a), t_h = \log(h / h_a), \\
 & t_x^* = (x^* - x_a) / w_a, t_y^* = (y^* - y_a) / h_a, t_w^* = \log(w^* / w_a), t_h^* = \log(h^* / h_a)
 \end{aligned} \tag{18}$$

[0057] 式中, x, y, w, h 分别表示包围框中心坐标、宽度和高度, x_a, y_a, w_a, h_a 分别表示候选区域中心坐标、宽度和高度, x^*, y^*, w^*, h^* 分别表示预测的包围框中心坐标、宽度和高度; 用公式 (18) 的计算结果进行位置精修; 实际上, 并没有显式地提取任何候选窗口, 完全使用区域建议网络自身完成判断和位置精修。

[0058] 所述的VGG网络, 在标签车辆图像数据上建立多层神经网络的方法, 分为两步, 一是每次训练一层网络, 二是调优, 使原始表示 X 向上生成的高级表示 r 和该高级表示 r 向下生成的 X' 尽可能一致;

[0059] 卷积神经网络的前向传播过程, 上一层的输出即为当前层的输入, 并通过激活函数逐层传递, 因此整个网络的实际计算输出用公式 (4) 表示,

$$[0060] \quad O_p = F_n(\dots(F_2(F_1(XW_1)W_2)\dots)W_n) \tag{4}$$

[0061] 式中, X 表示原始输入, F_1 表示第1层的激活函数, W_1 表示第1层的映射权值矩阵, O_p 表示整个网络的实际计算输出;

[0062] 当前层的输出用 (5) 表示,

$$[0063] \quad X^l = f^l(W^l X^{l-1} + b^l) \tag{5}$$

[0064] 式中, l 代表网络层数, X^l 表示当前层的输出, X^{l-1} 表示上一层的输出, 即当前层的输入, W^l 代表已经训练好的、当前网络层的映射权值矩阵, b^l 为当前网络的加性偏执, f^l 是当前网络层的激活函数; 采用的激活函数 f^l 为纠正线性单元, 即ReLU, 用公式 (6) 表示,

$$[0065] \quad f^l = \max((W^l)^T X^l, 0) = \begin{cases} (W^l)^T X^l & (W^l)^T X^l > 0 \\ 0 & (W^l)^T X^l \leq 0 \end{cases} \tag{6}$$

[0066] 式中, l 代表网络层数, W^l 代表已经训练好的、当前网络层的映射权值矩阵, f^l 是当前网络层的激活函数; 其作用是如果卷积计算结果小于0, 则让其为0; 否则保持其值不变。

[0067] 所述的VGG网络, 前5层是一个典型的深度卷积神经网络, 该神经网络训练是一个反向传播过程, 通过误差函数反向传播, 利用随机梯度下降法对卷积参数和偏置进行优化调整, 直到网络收敛或者达到最大迭代次数停止;

[0068] 反向传播需要通过对带有标签的训练样本进行比较, 采用平方误差代价函数, 对于 c 个类别, N 个训练样本的多类别进行识别, 网络最终输出误差函数用公式 (7) 来计算误差,

$$[0069] \quad E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (t_k^n - y_k^n)^2 \tag{7}$$

[0070] 式中, E^N 为平方误差代价函数, t_k^n 为第 n 个样本对应标签的第 k 维, y_k^n 为第 n 个样本对应网络预测的第 k 个输出;

[0071] 对误差函数进行反向传播时, 采用传统的BP算法类似的计算方法, 如公式 (8) 所示,

$$\begin{aligned}
 \delta^l &= (\mathbf{W}^{l+1})^T \delta^{l+1} \times f'(\mathbf{u}^l) \\
 \mathbf{u}^l &= \mathbf{W}^l \mathbf{x}^{l-1} + \mathbf{b}^l
 \end{aligned}
 \tag{8}$$

[0072] 式中, δ^l 代表当前层的误差函数, δ^{l+1} 代表上一层的误差函数, \mathbf{W}^{l+1} 为上一层映射矩阵, f' 表示激活函数的反函数, 即上采样, \mathbf{u}^l 表示未通过激活函数的上一层的输出, \mathbf{x}^{l-1} 表示下一层的输入, \mathbf{W}^l 为本层映射权值矩阵;

[0074] 通过误差反向传播后, 得到各个网络层的误差函数 δ^l , 然后采用随机梯度下降法对网络权值 \mathbf{W}^l 进行修改, 再进行下一次迭代, 直至网络达到收敛条件; 在进行误差传递时需要先通过公式 (8) 中的上采样使得前后两层尺寸相同, 再进行误差传递;

[0075] 算法思想是: 1) 首先逐层构建单层神经元, 这样每次都是训练一个单层网络; 2) 当所有层训练完后, 使用 wake-sleep 算法进行调优。

[0076] 深度学习训练过程具体如下:

[0077] STEP21: 使用自下而上的非监督学习, 即从底层开始, 一层一层的往顶层训练, 学习车辆图像特征: 先用无标签车辆图像数据训练第一层, 训练时先学习第一层的参数, 由于模型容量的限制以及稀疏性约束, 使得得到的模型能够学习到数据本身的结构, 从而得到比输入更具有表示能力的特征; 在学习得到第 1-1 层后, 将 1-1 层的输出作为第 1 层的输入, 训练第 1 层, 由此分别得到各层的参数; 具体计算如公式 (5)、(6) 所示;

[0078] STEP22: 自顶向下的监督学习, 即通过带标签的车辆图像数据去训练, 误差自顶向下传输, 对网络进行微调; 具体计算如公式 (7)、(8) 所示;

[0079] 基于 STEP21 得到的各层参数进一步微调整个多层模型的参数, 这一步是一个有监督训练过程; STEP21 类似神经网络的随机初始化初值过程, 由于深度学习的 STEP21 不是随机初始化, 而是通过学习输入数据的结构得到的, 因而这个初值更接近全局最优, 从而能够取得更好的效果。

[0080] 所述的 VGG 网络的前 5 层的模型初始化: 主要分为数据准备、计算图像均值、网络的定义、训练和恢复数据等 5 个步骤;

[0081] 1) 数据准备: 通过爬虫软件收集了各类车型的图像数据, 得到的基本上是带有标签的车辆图像数据, 将其作为训练图像数据; 另一类数据是通过卡口摄像机获取的车辆图像数据;

[0082] 2) 计算图像均值: 模型需要从每张图片减去均值;

[0083] 3) 网络的定义: 主要定义 xml 标签路径、图片的路径、存放 train.txt、val.txt、test.txt 和 trainval.txt 文件的路径;

[0084] 4) 训练: 运行训练模块;

[0085] 5) 恢复数据: 删除前 ReLU5 的层, 并将 roi_pool5 的 bottom 改为 data 和 rois;

[0086] 经过上述处理完成了车辆车型预训练的模型初始化工作。

[0087] 所述的区域建议网络利用了所述的 VGG 网络的前 5 层底层特征提取结果, 即两个网络共享了所述的 VGG 网络的前 5 层底层特征, 需要通过交替优化来学习优化共享的特征; 具体算法如下:

[0088] STEP41: 用区域建议网络的优化训练区域建议网络, 用上述的数据准备、计算图像均值、网络的定义、训练和恢复数据等 5 个步骤完成模型初始化, 并端到端微调用用于区域建议任务;

[0089] STEP42:用STEP1的区域建议网络生成的建议框,由快速R-CNN训练一个单独的检测网络,这个检测网络同样是由车辆车型预训练的模型初始化的,这时候两个网络还没有共享卷积层;

[0090] STEP43:用检测网络初始化区域建议网络训练,固定共享的卷积层,并且只微调区域建议网络独有的层,这时两个网络共享卷积层了;

[0091] STEP44:保持共享的卷积层固定,微调快速R-CNN的分类层;这样,两个网络共享相同的卷积层,最终构成一个统一的网络。

[0092] 本发明中,车辆车型视觉识别主要流程如下;

[0093] STEP51:读取待识别图像;

[0094] STEP52:对待识别图像进行归一化处理,得到RGB三个不同颜色 224×224 归一化的图像数据;

[0095] STEP53:将RGB三个不同颜色 224×224 归一化的图像数据输入到三个CNN通道,经5层卷积处理得到车辆车型特征图像数据;

[0096] STEP54:对车辆车型特征图像数据通过区域建议网络生成的建议框,选取一个最高得分的建议框,即得到一个感兴趣区域,RoI;将这个RoI通过最大池化5层进行处理得到一个 $6 \times 6 \times 256$ RoI的卷积图;

[0097] STEP55:将RoI的卷积图输出给两个同级的全连接的层处理后得到4096维的特征向量,作为softmax分类器的输入数据;

[0098] STEP56:对4096维的特征向量用softmax的分类回归分析得到车辆车型识别结果,识别出被测图像中的车辆属于哪一款车型。

[0099] 本发明的有益效果主要表现在:

[0100] 1) 提供了一种基于快速R-CNN深度神经网络的车型识别方法;

[0101] 2) 研发一种深度学习方法,实现无监督的车型特征提取;

[0102] 3) 设计一种适用于上千种的类型子类的分类器,并具有扩展性;

[0103] 4) 实现了一个真正意义上的使用一个快速R-CNN网络实现端到端的车辆检测和识别的框架,并具有适用于车辆目标的形态多样性,光照变化多样性,背景多样性等环境下快速、高精度和鲁棒性的车辆子类识别。

附图说明

[0104] 图1为边缘信息候选框的检测算法流程;

[0105] 图2为区域建议网络中的处理内容;

[0106] 图3为3种尺度和3种长宽比的滑动窗口示意图;

[0107] 图4为区域建议网络的概要图;

[0108] 图5为区域建议框生成说明图;

[0109] 图6为快速R-CNN网络中的共享网络的说明图;

[0110] 图7为对实际道路上车辆通过区域建议网络处理后得到的候选区域;

[0111] 图8为用代价函数实现区域建议网络自身完成判断和位置精修示意图;

[0112] 图9为基于CNN的车型识别流程框图;

[0113] 图10为wake-sleep算法说明图;

- [0114] 图11为VGG模型的CNN的车型识别流程框图；
[0115] 图12为快速R-CNN网络实现端到端的车辆检测和识别的训练过程图；
[0116] 图13为快速CNN网络实现端到端的车辆检测和识别流程概要图；
[0117] 图14为快速R-CNN网络实现端到端的车辆检测和识别的流程图。

具体实施方式

[0118] 下面结合附图对本发明的技术方案作进一步详细描述。

[0119] 实施例1

[0120] 参照图1~14,本发明解决其技术问题所采用的技术方案是:

[0121] (1)关于设计一种车辆对象的快速视觉分割算法;

[0122] 首先,设计一种车辆对象的快速视觉分割算法,即对车辆对象进行区域选择和定位;

[0123] 为了对车辆目标的位置进行定位;由于车辆目标可能出现在图像的任何位置,而且目标的大小、长宽比例也不确定,原有的技术是最初采用滑动窗口的策略对整幅图像进行遍历,而且需要设置不同的尺度,不同的长宽比;这种穷举的策略虽然包含了目标所有可能出现的位置,但是缺点也是显而易见的:时间复杂度太高,产生冗余窗口太多,这也严重影响后续特征提取和分类的速度和性能;

[0124] 对于滑动窗口存在的问题,本发明提出了一种候选区域的解决方案;即预先找出图中车辆目标可能出现的位置;由于候选区域利用了图像中的纹理、边缘、颜色等信息,能保证在选取较少窗口的情况下保持较高的召回率;这样能有效降低后续操作的时间复杂度,并且获取的候选窗口要比滑动窗口的质量更高;可选用的算法是选择性搜索,即selective Search和边缘信息候选框,即edge Boxes;这些算法的核心是利用了人类视觉的一眼“纵观全局”,直接发现车辆目标在整幅图像中的“大概的位置”;由于选择性搜索算法耗时比较大,不适用与在线的车型识别与检测;本发明采用边缘信息候选框的检测算法。

[0125] 边缘信息候选框的检测算法思想是:利用边缘信息,确定候选框内的轮廓个数和与候选框边缘重叠的轮廓个数,并基于此对候选框进行评分,进一步根据得分的高低顺序确定由大小,长宽比,位置构成的候选区域信息;边缘信息候选框的检测算法流程如图1所示;算法步骤如下:

[0126] STEP11:用结构深林边缘检测算法对原图像进行处理,得到的边缘图像,然后用非极大值抑制算法对边缘图像进一步处理得到一个相对稀疏的边缘图像;

[0127] STEP12:将相对稀疏的边缘图像中近乎在一条直线上的边缘点,集中起来形成一个边缘组,具体的做法是,不停地寻找8连通的边缘点,直到两两边缘点之间的方向角度差值的和大于 $\pi/2$,这样便得到了N多个边缘组 $s_i \in S$;

[0128] STEP13:用公式(1)计算两两边缘组之间的相似度,

[0129]
$$a(s_i, s_j) = |\cos(\theta_i - \theta_{ij}) \cos(\theta_j - \theta_{ij})|^\gamma \quad (1)$$

[0130] 式中, θ_i 和 θ_j 分别为两个边缘组的平均取向, s_i 和 s_j 分别表示两个边缘组, θ_{ij} 为两个边缘组的平均位置 x_i 和 x_j 之间的夹角, γ 为相似敏感系数,这里选择 $\gamma = 2$, $a(s_i, s_j)$ 两个边缘组之间的相似度;为了提高计算效率,这里将相似度 $a(s_i, s_j)$ 计算值超过阈值, $T_s \geq 0.05$ 的边缘组进行储存,其余均设置为零;

[0131] STEP14:给每一个边缘组赋予一个权值,权值计算方法由公式(2)给出,

$$[0132] \quad W_b(s_i) = 1 - \max_T \prod_j^{T-1} a(t_j, t_{j+1}) \quad (2)$$

[0133] 式中,T为从候选框的边缘开始到达 s_i 的边缘组序列集合的路径, $W_b(s_i)$ 为边缘 s_i 的权值, t_j 为……,(参数解释);如果没有找到路径就将 $W_b(s_i)$ 设定为1;

[0134] STEP15:用公式(3)计算候选框的评分,

$$[0135] \quad h_b = \frac{\sum_i W_b(s_i) m_i}{2(b_w + b_h)^k} \quad (3)$$

[0136] 式中, m_i 为对在边缘组 s_i 中所有边缘 p 的大小 m_p 的总和, $W_b(s_i)$ 为边缘 s_i 的权值, b_w 和 b_h 分别为候选框的宽度和高度, k 为大小系数,这里定义 $k=1.5$;计算窗口内边缘个数进行打分,最后排序打分来过滤掉低分的候选框;对于本发明主要应用于卡口车辆的提取,这里就选择最高分的候选框作为被测车辆对象前景图像;

[0137] (2)关于研发一种深度学习方法,实现无监督的车辆特征提取;

[0138] 由于车辆目标的形态多样性,光照变化多样性,背景多样性等因素使得设计一个鲁棒的特征并不是那么容易;然而提取特征的好坏直接影响到分类的准确性;

[0139] 满足上述三个多样性需求的、鲁棒的车辆特征提取必须通过无监督深度学习来实现完成,逐层初始化是一个非常有用的解决方案;深度学习的实质,是通过构建具有很多隐层的机器学习模型和海量的训练数据,来学习更有用的特征,从而最终提升分类或预测的准确性;因此,本发明中通过以下两点来实现深度学习:1)模型结构的深度有5层~10多层的隐层节点;2)通过逐层特征变换,将样本在原空间的特征表示变换到一个新特征空间,从而使分类或预测更加容易;

[0140] 本发明,提出了一种在非监督数据上,即无标签车辆图像数据上建立多层神经网络的方法,简单的说,分为两步,一是每次训练一层网络,二是调优,使原始表示 X 向上生成的高级表示 r 和该高级表示 r 向下生成的 X' 尽可能一致;

[0141] 卷积神经网络的前向传播过程,上一层的输出即为当前层的输入,并通过激活函数逐层传递,因此整个网络的实际计算输出用公式(4)表示,

$$[0142] \quad O_p = F_n(\dots(F_2(F_1(XW_1)W_2)\dots)W_n) \quad (4)$$

[0143] 式中, X 表示原始输入, F_1 表示第1层的激活函数, W_1 表示第1层的映射权值矩阵, $l=1,2,3,\dots$ 代表网络层数, O_p 表示整个网络的实际计算输出;

[0144] 当前层的输出用公式(5)表示,

$$[0145] \quad X^l = f^l(W^l X^{l-1} + b^l) \quad (5)$$

[0146] 式中, l 代表网络层数, X^l 表示当前层的输出, X^{l-1} 表示上一层的输出,即当前层的输入, W^l 代表已经训练好的、当前网络层的映射权值矩阵, b^l 为当前网络的加性偏执, f^l 是当前网络层的激活函数;本发明采用的激活函数 f^l 为纠正线性单元,即Rectified Linear Units,ReLU,用公式(6)表示,

$$[0147] \quad f^l = \max((W^l)^T X^l, 0) = \begin{cases} (W^l)^T X^l & (W^l)^T X^l > 0 \\ 0 & (W^l)^T X^l \leq 0 \end{cases} \quad (6)$$

[0148] 式中, l 代表网络层数, W^l 代表已经训练好的、当前网络层的映射权值矩阵, f^l 是当前网络层的激活函数;其作用是如果卷积计算结果小于0,则让其为0;否则保持其值不变;

[0149] 在CNN中采用局部感受野以及权值共享的方法来进一步降低网络参数。所谓局部感受野,指的是每种卷积核只与图像中的某个特定区域相连接,即每种卷积核只卷积图像的一部分,接下来再在其他层中将这局部卷积特征联系在一起,既符合图像像素在空间上的关联性,又降低了卷积参数数量。而权值共享则是使得每种卷积核的权值均相同,通过增加卷积核的种类来提取图像多方面特征;为了给车辆车型子类分类提供更多的细节特征,本发明中适当地增加了卷积核的种类;

[0150] 卷积神经网络训练是一个反向传播过程,与传统的BP算法类似,通过误差函数反向传播,利用随机梯度下降法对卷积参数和偏置进行优化调整,直到网络收敛或者达到最大迭代次数停止。

[0151] 反向传播需要通过对带有标签的训练样本进行比较,计算误差。例如采用平方误差代价函数,对于 c 个类别, N 个训练样本的多类别识别问题,网络最终输出误差函数用公式 (7) 表示,

$$[0152] \quad E^N = \frac{1}{2} \sum_{n=1}^N \sum_{k=1}^c (\mathbf{t}_k^n - \mathbf{y}_k^n)^2 \quad (7)$$

[0153] 式中, E^N 为平方误差代价函数, \mathbf{t}_k^n 为第 n 个样本对应标签的第 k 维, \mathbf{y}_k^n 为第 n 个样本对应网络预测的第 k 个输出;

[0154] 对误差函数进行反向传播时,采用传统的BP算法类似的计算方法,如公式 (8) 所示,

$$[0155] \quad \begin{aligned} \delta^l &= (\mathbf{W}^{l+1})^T \delta^{l+1} \times f'(\mathbf{u}^l) \\ \mathbf{u}^l &= \mathbf{W}^l \mathbf{x}^{l-1} + \mathbf{b}^l \end{aligned} \quad (8)$$

[0156] 式中, δ^l 代表当前层的误差函数, δ^{l+1} 代表上一层的误差函数, W^{l+1} 为上一层映射矩阵, f' 表示激活函数的反函数,即上采样, u^l 表示未通过激活函数的上一层的输出, x^{l-1} 表示下一层的输入, W^l 为本层映射权值矩阵;

[0157] 通过误差反向传播后,得到各个网络层的误差函数 δ^l , 然后采用随机梯度下降法对网络权值 W^l 进行修改,再进行下一次迭代,直至网络达到收敛条件;要注意的是,由于层与层之间的大小尺寸不同,在进行误差传递时需要先通过公式 (8) 中的上采样使得前后两层尺寸相同,再进行误差传递;

[0158] 算法思想是:1) 首先逐层构建单层神经元,这样每次都是训练一个单层网络;2) 当所有层训练完后,使用wake-sleep算法进行调优。

[0159] 深度学习训练过程具体如下:

[0160] STEP21: 使用自下而上的非监督学习,即从底层开始,一层一层的往顶层训练,学习车辆图像特征:先用无标签车辆图像数据训练第一层,训练时先学习第一层的参数,由于模型容量的限制以及稀疏性约束,使得得到的模型能够学习到数据本身的结构,从而得到比输入更具有表示能力的特征;在学习得到第1-1层后,将1-1层的输出作为第1层的输入,训练第1层,由此分别得到各层的参数;具体计算如公式 (5)、(6) 所示;

[0161] STEP22: 自顶向下的监督学习,即通过带标签的车辆图像数据去训练,误差自顶向

下传输,对网络进行微调:具体计算如公式(7)、(8)所示;

[0162] 基于STEP21得到的各层参数进一步微调整个多层模型的参数,这一步是一个有监督训练过程;STEP21类似神经网络的随机初始化初值过程,由于深度学习的STEP21不是随机初始化,而是通过学习输入数据的结构得到的,因而这个初值更接近全局最优,从而能够取得更好的效果;所以深度学习效果好坏很大程度上归功于STEP21的特征学习过程;

[0163] 对于带标签的车辆图像数据集,本发明采用网络爬虫技术收集各种车型的图像,这样收集到的车辆图像再经过人工确认作为带标签的车辆图像数据;

[0164] 关于wake-sleep算法,见附图8,其主要思想是在wake阶段对给定生成权值,通过学习得到认知权值;在sleep阶段对给定认知权值,通过学习得到生成权值;

[0165] wake阶段,第1层生成权值 g_1 ,用公式(9)进行更新,

$$[0166] \quad \Delta g_1 = \varepsilon s^{1+1} (s^1 - p^1) \quad (9)$$

[0167] 式中, Δg_1 为第1层生成权值 g_1 的更新变化值, ε 为学习速率, s^{1+1} 为第1+1层神经元活跃度, s^1 为第1层神经元活跃度, p^1 为第1层神经元的当前状态驱动时的激活概率;

[0168] sleep阶段,第1层认知权值 w_1 ,用公式(10)进行更新,

$$[0169] \quad \Delta w_1 = \varepsilon s^{1-1} (s^1 - q^1) \quad (10)$$

[0170] 式中, Δw_1 为第1层认知权值 w_1 的更新变化值, ε 为学习速率, s^{1-1} 为第1-1层神经元活跃度, s^1 为第1层神经元活跃度, q^1 为第1层神经元被使用当前认知权值的前一层神经元的当前状态驱动时的激活概率;

[0171] (3)设计一种适用于上千种的类型子类的分类器,并具有扩展性;

[0172] 本发明采用Softmax分类器对车型进行分类;Softmax分类器在特征鲁棒的情况下,具有较好的分类效果,同时这种分类器具有扩展性,在新的车型出现后无需对原有训练好的网络特征进行重新学习训练,增加了系统的实用性;softmax原理是将提取到的输入数据特征与特征库进行比对,计算出每一个分类结果的概率,然后取概率最高的一个结果进行输出;

[0173] 将CNN中的学习结果作为softmax分类器的输入数据;Softmax回归是面向多类分类问题的Logistic回归,是Logistic回归的一般形式,适用于类别之间互斥的情况。假设对于训练集 $\{(x^{(1)}, y^{(1)}), \dots, (x^{(m)}, y^{(m)})\}$,有 $y^{(i)} \in \{1, 2, \dots, k\}$,对于给定的样本输入 x ,输出一个 k 维的向量来表示每一种分类结果出现的概率为 $p(y=i|x)$,假设函数 $h(x)$ 如下:

$$[0174] \quad h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}, \theta) \\ p(y^{(i)} = 2 | x^{(i)}, \theta) \\ \vdots \\ p(y^{(i)} = k | x^{(i)}, \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^k e^{\theta_j^T x^{(i)}}} \begin{bmatrix} e^{\theta_1^T x^{(i)}} \\ e^{\theta_2^T x^{(i)}} \\ \vdots \\ e^{\theta_k^T x^{(i)}} \end{bmatrix} \quad (11)$$

[0175] $\theta_1, \theta_2, \dots, \theta_k$ 是模型的参数,并且所有的概率和为1。加入规则项后的代价函数为:

$$[0176] \quad J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m \sum_{j=1}^k 1\{y^{(i)} = j\} \log \frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^k e^{\theta_l^T x^{(i)}}} \right] + \frac{\lambda}{2} \sum_{i=1}^k \sum_{j=0}^n \theta_{ij}^2 \quad (12)$$

[0177] 代价函数对第 j 个类别的第1个参数的偏导数为:

$$[0178] \quad \nabla_{\theta_j} J(\theta) = -\frac{1}{m} \sum_{i=1}^m \left[x^{(i)} (\{y^{(i)} = j\} - p(y^{(i)} = j | x^{(i)}; \theta)) \right] + \lambda \theta_j \quad (13)$$

[0179] 最后,通过最小化 $J(\theta)$,实现softmax的分类回归。

[0180] (4) 设计一个真正意义上的使用一个快速R-CNN网络实现端到端的车辆检测和识别的框架:

[0181] 本发明要着重解决以下三个问题:

[0182] 1) 如何设计区域建议网络;

[0183] 2) 如何训练区域建议网络;

[0184] 3) 如何让区域建议网络和快速R-CNN网络共享特征提取网络;

[0185] 区域建议网络内容主要包括:区域建议网络结构、特征提取、区域建议网络的设计和训练思路、候选区域、窗口分类和位置精修;

[0186] 区域建议网络结构、特征提取、区域建议网络的设计:

[0187] 区域建议网络结构、特征提取、区域建议网络的设计如附图11所示,为了能在多个GPU上进行计算,在层内计算中划分为若干组,每组中的计算由其对应GPU完成,这样能提升计算速度;图12所示是一种VGG网络,即visual geometry group,8个卷基层,3个全连接层,共计11层;8个卷基层中有5个组的卷积层、2个分类层提取图像特征、1个分类层分类特征;3个全连接层分别连接分类层6、分类层7和分类层8;

[0188] 图11中,归一化的 224×224 图像直接送入网络,前五阶段是基础的卷积+ReLU+池化的形式,在第五阶段结尾处再输入 p 个候选区域,候选区域带有1个图像序号和4个几何位置信息;在第五阶段结尾处的RoI-池化层将每个候选区域均匀分成 $M \times N$ 块,对每块进行最大池化操作;将特征图上大小不一的候选区域转变为大小统一的数据,送入下一层训练和识别;上述前五阶段技术都是卷积神经网络技术中被证明成熟的技术,关键是如何使得这些候选区域可以复用这些图片前五个阶段的网络特征;

[0189] 区域建议网络和快速R-CNN网络共享特征提取网络:

[0190] 附图6描述了区域建议网络和快速R-CNN如何共享的卷积层输出,图11中前五阶段属于一个原始特征提取网络,之后分别给区域建议网络生成区域建议框和给快速R-CNN检测区域建议框内的图像特征,然后输出给两个同级的全连接的层,即图11中的分类层6+ReLU6和分类层7+ReLU7,得到 p 个含有4096向量的特征图,接着分别送给分类层和窗口回归层进行处理;

[0191] 生成区域建议框:

[0192] 为了生成区域建议框,本发明在最后一个共享的卷积层输出的卷积特征映射上滑动小网络,这个网络全连接到输入卷积特征映射的 $n \times n$ 的空间窗口上,如附图5所示;每个滑动窗口映射到一个低维向量上,附图5中的低维向量为 $256-d$,每个特征映射的一个滑动窗口对应一个数值;这个向量输出给两个同级的全连接的层;一个窗口回归层和一个分类层;窗口回归层输出每一个位置上,9种推荐区域对应窗口需要具有平移缩放不变性,回归层从256维特征中输出4个平移缩放参数;分类层从256维特征中输出属于前景和背景的概率;

[0193] 比如,在每一个滑动窗口的位置,同时预测 k 个区域建议,因此窗口回归层有 $4k$ 个输出,即 k 个建议框的坐标编码;分类层输出 $2k$ 个建议框得分,即对每个建议框是车辆目标/

非车辆目标的估计概率；

[0194] 候选区域:k个建议框被相应的k个称为候选区域的框参数化,即每个候选区域以当前滑动窗口中心为中心,并对应一种尺度和长宽比,本发明使用3种尺度和3种长宽比,如附图3所示;这样在每一个滑动位置就有k=9种候选区域;对于大小为W×H的卷积特征映射,总共有W×H×k个候选区域;

[0195] 在整个快速R-CNN算法中,共有三种图像尺度:1)原图尺度:原始输入图像的大小,不受任何限制,不影响性能;2)归一化尺度:输入特征提取网络的大小,在测试时设置,候选区域在这个尺度上设定;这个参数和候选区域的相对大小决定了想要检测的目标范围;3)网络输入尺度:输入特征检测网络的大小,在训练时设置,为224×224。

[0196] 综上所述,区域建议网络是将一个图像作为输入,输出矩形目标建议框的集合,每个框有一个车辆对象的得分,如附图7所示;

[0197] 训练区域建议网络内容主要包括:训练样本、代价函数和超参数;

[0198] 区域建议网络的训练:

[0199] 区域建议网络的训练,本发明给每个候选区域分配一个二进制的标签,即是否是车辆对象;这里分配正标签给两类候选区域:(i)与某个GT,ground truth的包围区域有最高的交集并集之比,IoU,Intersection-over-Union,重叠的候选区域;(ii)与任意GT包围区域大于0.7的IoU交叠的候选区域;同时分配负标签给与所有GT包围区域的IoU比率都低于0.3的候选区域;删去非正非负的候选区域;

[0200] 训练样本算法:

[0201] STEP31:顺序读取训练集中的每张图;

[0202] STEP32:对每个标定的真值候选区域,与其重叠比例最大的候选区域记为前景样本;

[0203] STEP33:对STEP32剩余的候选区域,如果其与某个标定重叠,IoU比例大于0.7,记为前景样本;如果其与任意一个标定的重叠比例都小于0.3,记为背景样本;

[0204] STEP34:对STEP32和STEP33剩余的候选区域弃去不用;

[0205] STEP35:对跨越图像边界的候选区域弃去不用。

[0206] 代价函数:

[0207] 根据这些定义,这里遵循的多任务代价,采用最小化目标函数;对一个图像的代价函数用公式(14)表示,

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (14)$$

[0209] 式中,i是一个批处理中候选区域的索引, N_{cls} 为分类层的归一化系数, N_{reg} 为回归层的归一化系数, λ 为平衡权重, p_i 为车辆目标的预测概率, p_i^* 为GT标签,如果候选区域为正 $p_i^* = 1$,如果候选区域为负 $p_i^* = 0$; t_i 为一个向量,表示预测的包围框的4个参数化坐标, t_i^* 为与正候选区域对应的GT包围框的坐标向量, L_{cls} 为分类的对数代价, L_{reg} 为回归对数代价, $L(\{p_i\}, \{t_i\})$ 为总的对数代价;

[0210] 分类的对数代价 L_{cls} 由公式(15)计算,

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1 - p_i^*)(1 - p_i)] \quad (15)$$

[0212] 窗口回归对数代价 L_{reg} 由公式(16)计算,

$$[0213] \quad L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*) = R(\mathbf{t}_i - \mathbf{t}_i^*) \quad (16)$$

[0214] 式中, R 为定义的鲁棒的代价函数,属于Smooth L1误差,对离群值不敏感,用公式(17)计算,

$$[0215] \quad R(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{其他} \end{cases} \quad (17)$$

[0216] 公式(14)中的 $p_i^* L_{reg}(\mathbf{t}_i, \mathbf{t}_i^*)$ 这一项意味着只有正候选区域,即 $p_i^* = 1$ 时才有回归代价,其他情况由于 $p_i^* = 0$ 没有回归代价;分类层和窗口回归层的输出分别由 $\{p_i\}$ 和 $\{t_i\}$ 组成,这两项分别由 N_{cls} 和 N_{reg} 以及一个平衡权重 λ 归一化,这里选择 $\lambda = 10$, $N_{cls} = 256$, $N_{reg} = 2400$,通过这样的选择分类层和窗口回归层项差不多是等权重的;

[0217] 关于位置精修,这里采用4个值,中心坐标、宽度和高度,计算方法如下,

$$[0218] \quad \begin{aligned} t_x &= (x - x_a) / w_a, t_y = (y - y_a) / h_a, t_w = \log(w / w_a), t_h = \log(h / h_a), \\ t_x^* &= (x^* - x_a) / w_a, t_y^* = (y^* - y_a) / h_a, t_w^* = \log(w^* / w_a), t_h^* = \log(h^* / h_a) \end{aligned} \quad (18)$$

[0219] 式中, x 、 y 、 w 、 h 分别表示包围框中心坐标、宽度和高度, x_a 、 y_a 、 w_a 、 h_a 分别表示候选区域中心坐标、宽度和高度, x^* 、 y^* 、 w^* 、 h^* 分别表示预测的包围框中心坐标、宽度和高度;用公式(18)的计算结果进行位置精修;实际上,并没有显式地提取任何候选窗口,完全使用区域建议网络自身完成判断和位置精修;

[0220] 区域建议网络的优化:

[0221] 区域建议网络能自然地实现为全卷积网络,通过反向传播和随机梯度下降端到端训练;这里采用图像中心的采样策略训练这个网络,每个批处理由包含了许多正负样本的单个图像组成;随机地在一个图像中采样256个候选区域,用公式(14)计算批处理中的代价函数,其中采样的正负候选区域的比例是1:1;如果一个图像中的正样本数小于128,我们就用负样本填补这个批处理中的其他剩余候选区域;这里将批处理的大小设置为256;

[0222] 通过从零均值标准差为0.01的高斯分布中获取的权重来随机初始化所有新层,所谓新层指的是在区域建议网络后面的层,如附图11中的分类层6+ReLU6和分类层7+ReLU7;所有其他层,即共享的卷积层,如附图11中的前五层,通过对车辆车型的分类样本预训练的模型来初始化的;本发明在车辆车型数据集上对于60k个批处理用的学习率为0.001,对于下一20k个批处理用的学习率是0.0001;动量是0.9,权重衰减为0.0005;

[0223] 车辆车型预训练的模型初始化:主要分为数据准备、计算图像均值、网络的定义、训练和恢复数据等5个步骤;

[0224] 1) 数据准备;在data中新建文件夹myself,我们通过爬虫软件收集了各类车型的图像数据,由于带了关键字进行搜索得到的基本上是带有标签的车辆图像数据,我们将其作为训练数据;另一类数据是我们通过卡口摄像机获取的车辆图像数据;

[0225] 训练和测试的输入是用train.txt和val.txt描述的,列出所有文件和他们的标签;分类的名字是ASCII码的顺序,即0-999,对应的分类名和数字的映射在synset_

words.txt中;val.txt不能标签,全部设置成0;然后将图片的大小统一成 256×256 ;然后在caffe-master/examples中新建myself文件夹,然后将caffe-master/examples/imagenet的create_imagenet.sh复制到该文件夹下,将其名改为create_animal.sh,修改训练和测试路径的设置,运行该sh;最后得到myself_train_lmdb和myself_val_lmdb;

[0226] 2) 计算图像均值;模型需要我们从每张图片减去均值,所以我们必须获得训练的均值,用tools/compute_image_mean.cpp实现,同样复制caffe-master/examples/imagenet的./make_imagenet_mean到examples/myself中,将其改名为make_car_mean.sh,加以修改路径;

[0227] 3) 网络的定义;把caffe-master/models/bvlc_reference_caffenet中所有文件复制到caffe-master/examples/myself文件夹中,修改train_val.prototxt,注意修改数据层的路径;

[0228] 在训练中,我们用一个softmax—loss层计算损失函数和初始化反向传播,而在验证,我们使用精度层检测我们的精度;还有一个运行的协议solver.prototxt,复制过来,将第一行路径改为我们的路径net:“examples/myself/train_val.prototxt”,

[0229] test_iter:1000是指测试的批次;test_interval:1000是指每1000次迭代测试一次;base_lr:0.01是基础学习率;lr_policy:“step”学习率变化;gamma:0.1学习率变化的比率;stepsize:100000每100000次迭代减少学习率;display:20每20层显示一次;max_iter:450000最大迭代次数;momentum:0.9学习的参数;weight_decay:0.0005学习的参数;snapshot:10000每迭代10000次显示状态;solver_mode:GPU末尾加一行,代表用GPU运算;

[0230] 4) 训练;把caffe-master/examples/imagenet中的train_caffenet.sh复制过来并修改名为train_myself.sh运行,修改里面的路径;

[0231] 5) 恢复数据;把caffe-master/examples/imagenet中的resume_training.sh复制过来并运行;

[0232] 经过上述处理完成了车辆车型预训练的模型初始化工作;进一步,本发明提出一种4步训练算法,通过交替优化来学习共享的特征;

[0233] STEP41:用区域建议网络的优化训练区域建议网络,用上述数据准备、计算图像均值、网络的定义、训练和恢复数据等5个步骤完成模型初始化,并端到端微调用子区域建议任务;

[0234] STEP42:用STEP1的区域建议网络生成的建议框,由快速R-CNN训练一个单独的检测网络,这个检测网络同样是由车辆车型预训练的模型初始化的,这时候两个网络还没有共享卷积层;

[0235] STEP43:用检测网络初始化区域建议网络训练,固定共享的卷积层,并且只微调区域建议网络独有的层,这时两个网络共享卷积层了;

[0236] STEP44:保持共享的卷积层固定,微调快速R-CNN的分类层;这样,两个网络共享相同的卷积层,最终构成一个统一的网络;

[0237] 车辆车型的视觉识别:

[0238] 下面我们给出一个车辆车型视觉识别主要流程,整个处理流程如附图13所示;

[0239] STEP51:读取待识别图像;

[0240] STEP52:对待识别图像进行归一化处理,得到RGB三个不同颜色 224×224 归一化的

图像数据；

[0241] STEP53:将RGB三个不同颜色 224×224 归一化的图像数据输入到三个CNN通道,经5层卷积处理得到车辆车型特征图像数据；

[0242] STEP54:对车辆车型特征图像数据通过区域建议网络生成的建议框,选取一个最高得分的建议框,即得到一个感兴趣区域,RoI;将这个RoI通过最大池化5层进行处理得到一个 $6 \times 6 \times 256$ RoI的卷积图；

[0243] STEP55:将RoI的卷积图输出给两个同级的全连接的层处理后得到4096维的特征向量,作为softmax分类器的输入数据；

[0244] STEP56:对4096维的特征向量用softmax的分类回归分析得到车辆车型识别结果,本发明中将车辆车型分类成1000种类,从而识别出被测图像中的车辆属于哪一款车型。

[0245] 实施例2

[0246] 本发明的视觉识别技术具有普适性,适用于其他对象的子类识别,只要参与训练的数据运行在本发明开发的系统中进行学习,获得该类对象的特征后就能实现该类对象的子类识别任务。

[0247] 实施例3

[0248] 本发明的视觉识别技术具有扩展性,在新的子类出现后无需对原有训练好的网络特征进行重新学习训练,只要对新的子类进行训练学习,并且在系统中的softmax分类器扩展分类的数据即可。

[0249] 以上所述仅为本发明的较佳实施举例,并不用于限制本发明,凡在本发明精神和原则之内,所作的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

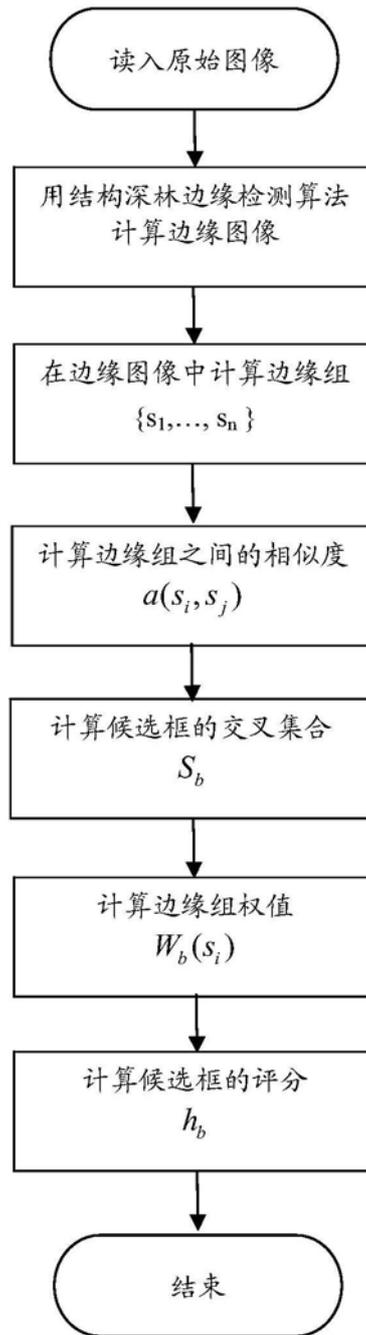


图1



图2

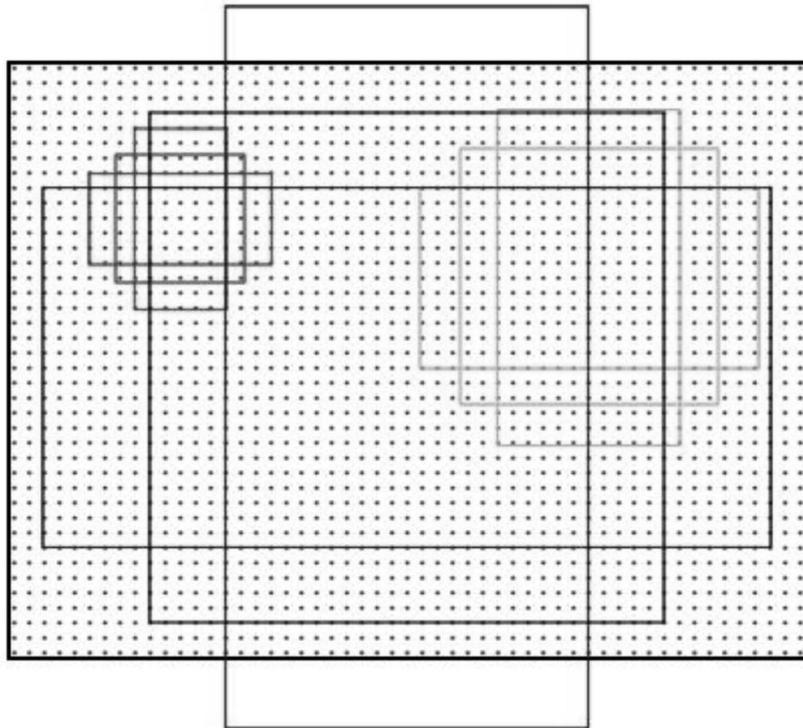


图3

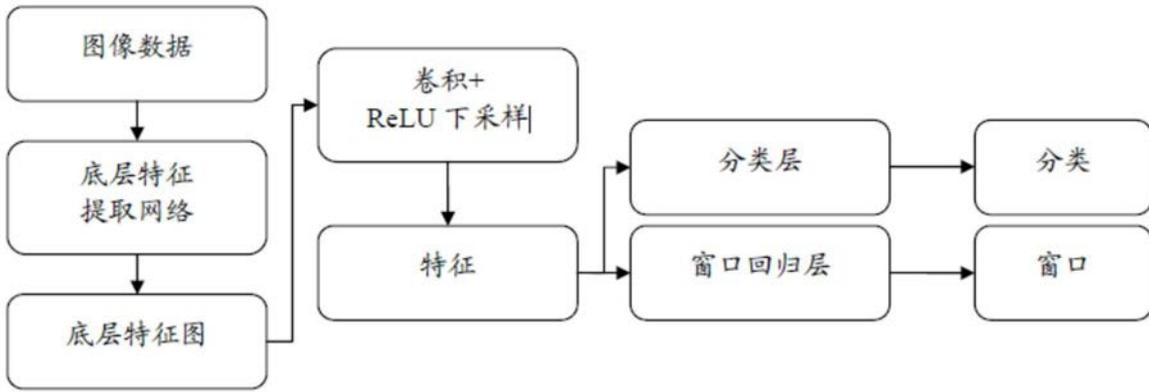


图4

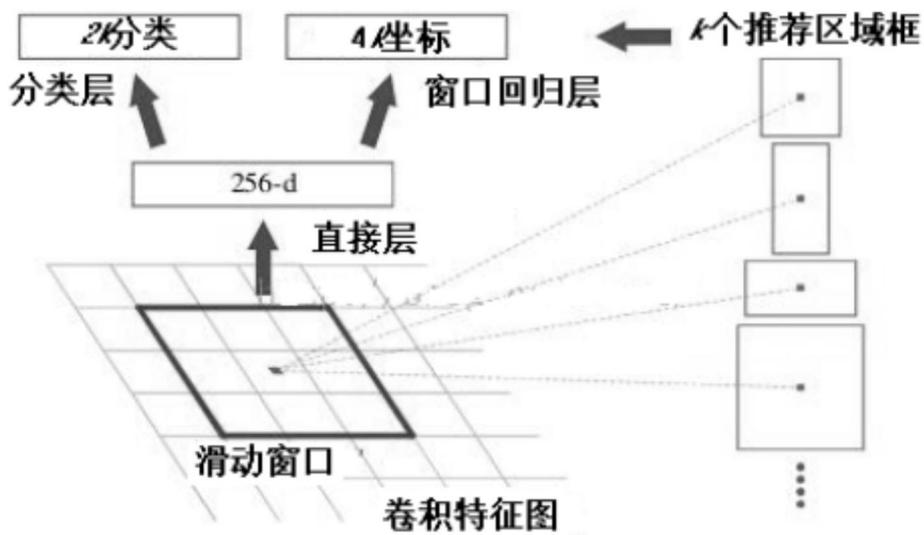


图5

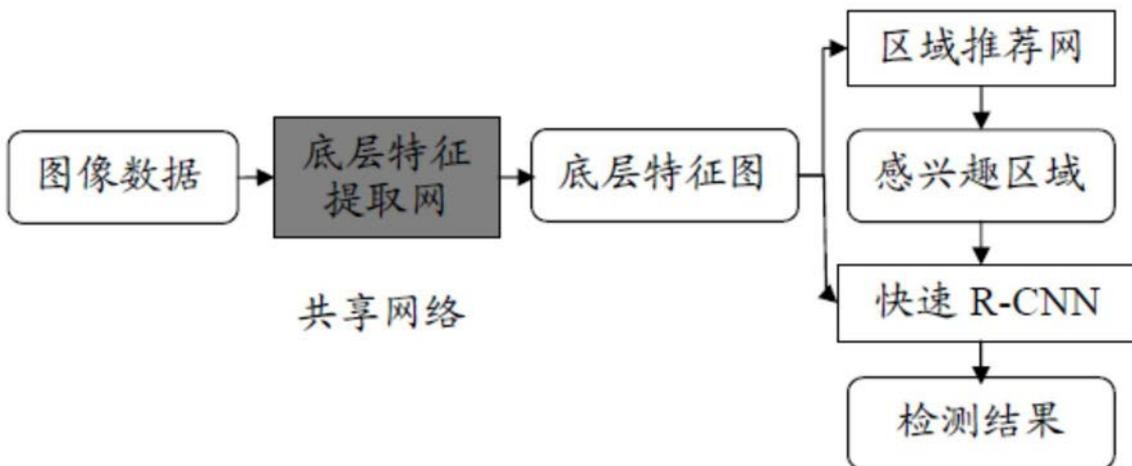


图6



图7

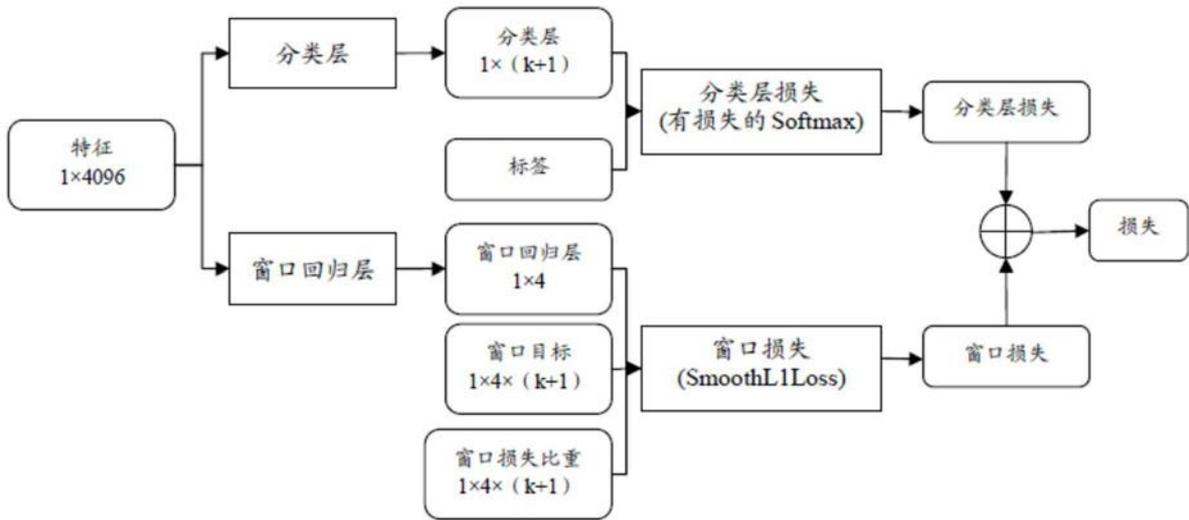


图8

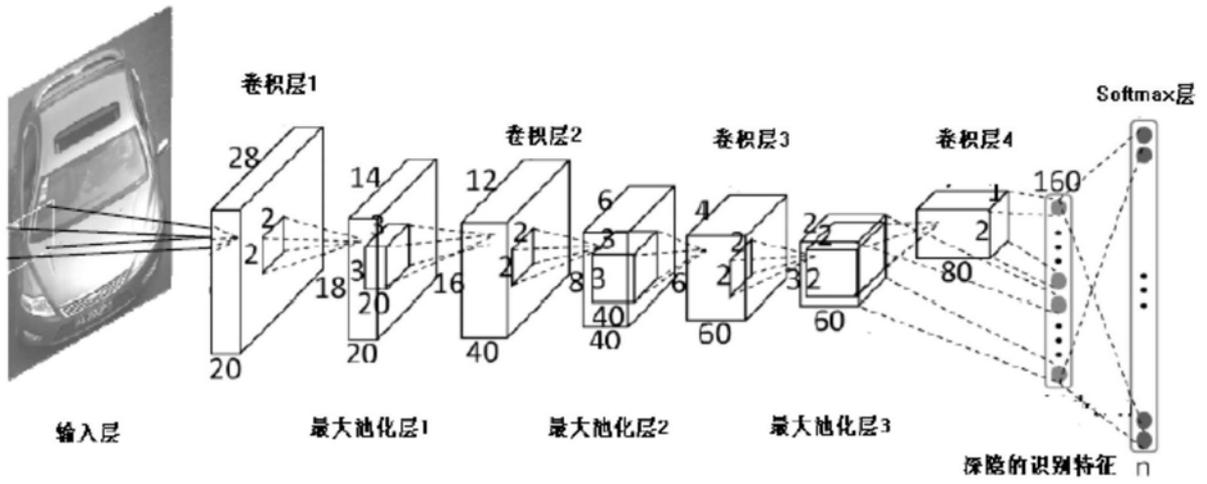


图9

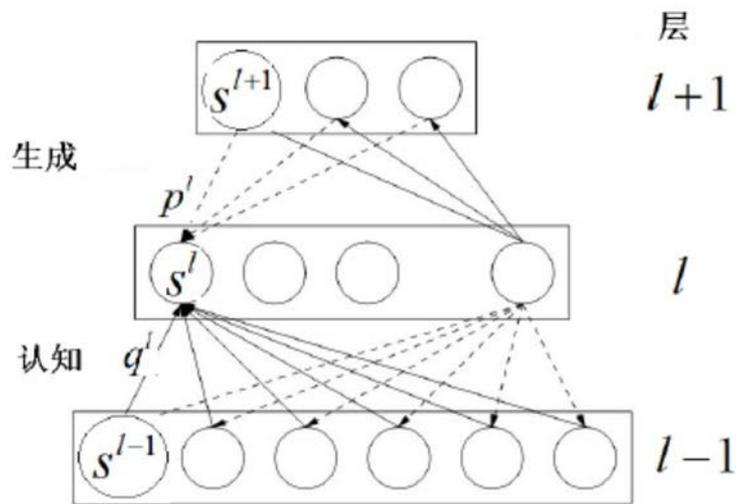


图10

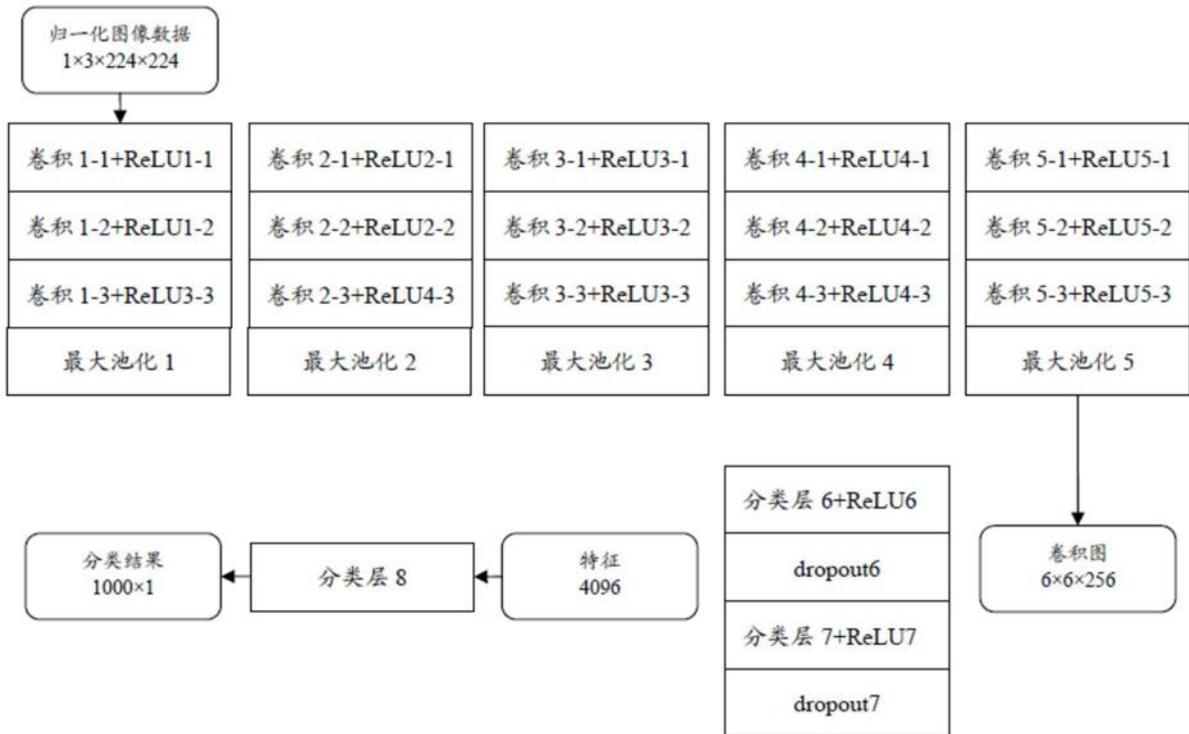


图11

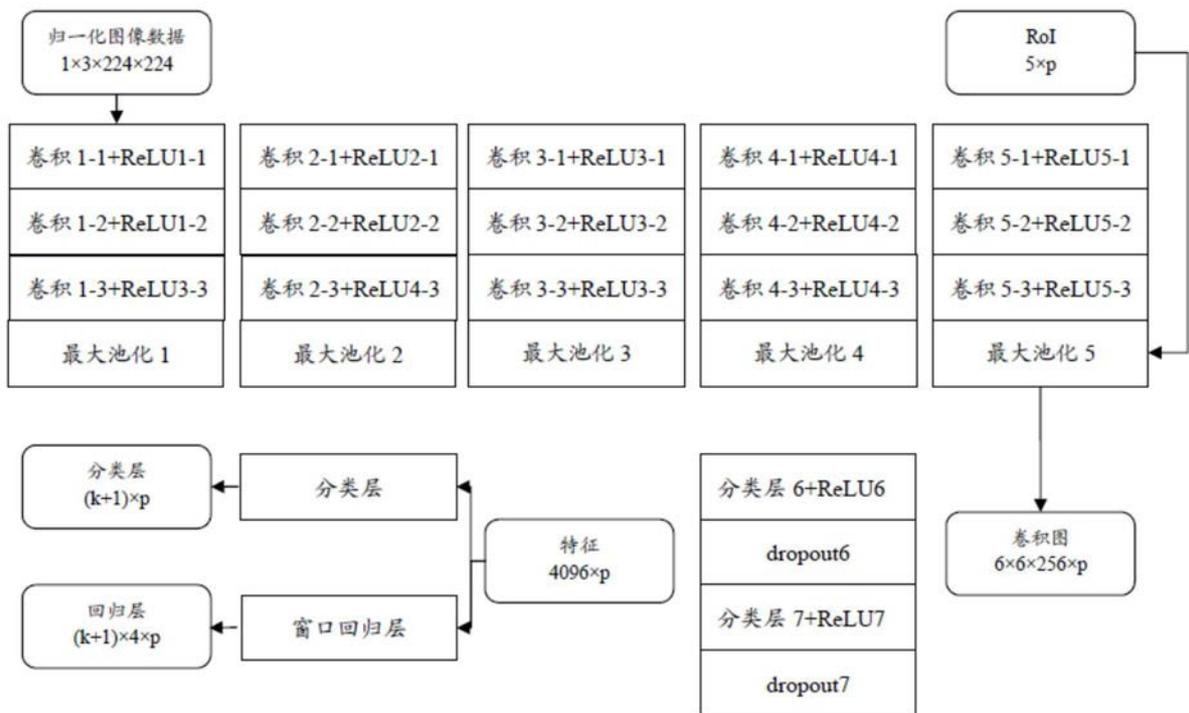


图12

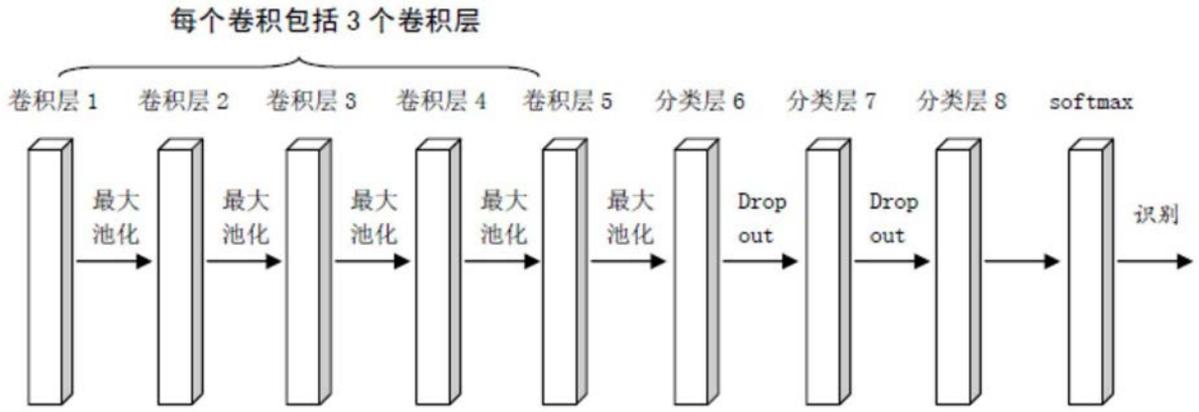


图13

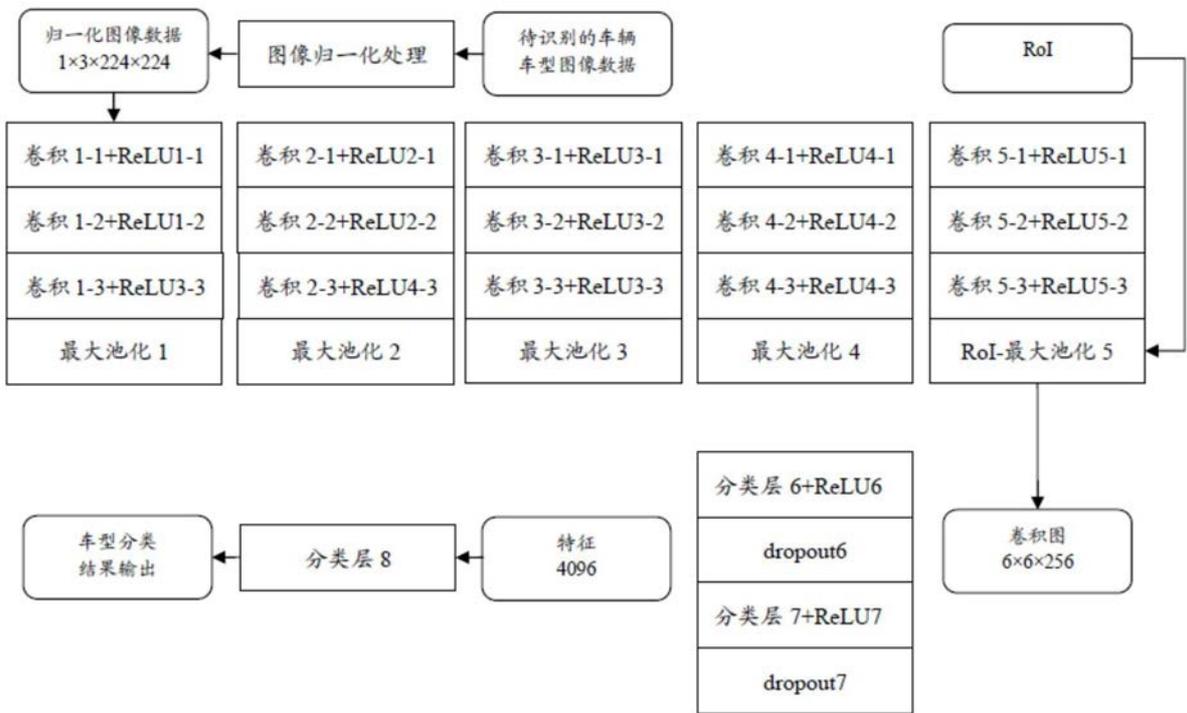


图14