

# 基于深度学习的目标检测与可行域分割研究

李立凯<sup>1</sup>, 卢焯华<sup>1,2</sup>, 邹斌<sup>1,2\*</sup>

<sup>1</sup> 武汉理工大学现代汽车零部件技术湖北省重点实验室, 湖北 武汉 430070;

<sup>2</sup> 汽车零部件技术湖北省协同创新中心, 湖北 武汉 430070

**摘要** 为提高智能车快速检测物体对多种场景的适应能力, 提出了多任务共享一个特征提取网络的联合方法。首先用 ResNet-50 作为编码器提取图像的特征; 然后采用单发多框检测算法的多尺度特征预测和快速回归思想, 对检测结果进行解码, 采用 DeepLab v3 中的多孔空间金字塔池化结构, 对经 ResNet-50 下采样后的图像特征进行多尺度映射、双线性上采样和批次归一化处理, 完成分割解码; 最后设定好参数训练联合方法。实验结果表明, 该方法的平均精度均值为 89.00%, 分割平均交并比为 83.0, 每秒传输帧数为 31 frame, 满足智能车的应用需求。

**关键词** 图像处理; 共享网络; 多尺度特征; 多任务联合; 深度学习

中图分类号 TP391

文献标志码 A

doi: 10.3788/LOP57.121013

## Research on Target Detection and Feasible Region Segmentation Based on Deep Learning

Li Likai<sup>1</sup>, Lu Chihua<sup>1,2</sup>, Zou Bin<sup>1,2\*</sup>

<sup>1</sup> Hubei Key Laboratory of Advanced Technology for Automotive Components, Wuhan University of Technology, Wuhan, Hubei 430070, China;

<sup>2</sup> Hubei Collaborative Innovation Center for Automotive Components Technology, Wuhan, Hubei 430070, China

**Abstract** In order to improve the adaptability of intelligent vehicles to quickly detect objects in various scenes, a joint method of multi-task sharing the same feature extraction network is proposed. First, ResNet-50 network is used to extract image features of the encoder. Then, multi-scale feature prediction and fast regression in single shot multibox detector target detection algorithm are used to decode the detection results. A pyramid pool structure of porous space in DeepLab v3 is used to process the multi-scale mapping, bilinear sampling and batch normalization of the image features after ResNet-50 sampling so as to complete segmentation and decoding. Finally, the training of the joint method is completed under the set training parameters. Experimental results show that the mean average precision of the method is 89.00%, the mean intersection over union is 83.0, and the number of frames per second is 31 frame, which can support intelligent vehicle to complete certain tasks.

**Key words** image processing; shared network; multi-scale feature; multi-task coalition; deep learning

**OCIS codes** 100.4996; 100.3008

## 1 引言

近年来, 无人驾驶车辆视觉感知技术的研究促进了智能车的发展。视觉感知系统除了要准确识别目标外, 还需具有较强的实时性。因此设计准确性和实时性满足无人驾驶车辆要求的识别和分割方法, 可为后续车辆的避碰和路径规划等任务提供重要支撑。

由于基于深度学习的目标检测和分割模型实时性、准确性、鲁棒性较好, 在无人驾驶视觉感知系统中得到了广泛的应用。随着卷积神经网络(CNN)层数的逐渐加深, 网络的特征提取、抗尺度变化、抽象和抗平移等能力也逐渐变强<sup>[1]</sup>。赵恒等<sup>[2]</sup>基于改进的跨层级特征融合循环全卷积神经网络, 提出了一种结合深度学习的图像显著目标检测算法, 有效提高了复杂场景图像的目标检测精度; 徐

收稿日期: 2019-09-05; 修回日期: 2019-10-03; 录用日期: 2019-11-02

\* E-mail: zbegn@qq.com

乐先等<sup>[3]</sup>利用 TensorFlow 深度学习平台结合单发多框检测器(SSD)算法,有效识别了车位分布及空车位;张祥甫等<sup>[4]</sup>综述了基于深度学习的语义分割问题,探讨了 CNN 取得的突破性进展,阐述了深度学习方法处理语义分割问题的具大潜力。Shelhamer 等<sup>[5]</sup>提出的全卷积网络(FCN)算法,在精度和速度上的表现超越了以机器学习为主的传统分割算法,表明了 CNN 在分割任务中具有较强的泛化能力。基于 FCN 的 SegNet<sup>[6]</sup>、DeepLab<sup>[7]</sup>等算法也都取得了一定的成绩,其中 DeepLab 算法经过三代网络架构和深度信息的优化及其简单高效的解码网络,大大提高了语义的分割精度,但对复杂场景图像的目标分割实时性仍有待提高。综上所述,基于深度学习的目标检测和分割算法在多个领域都有所突破,但网络训练需要的计算资源较多,如何在计算资源有限的情况下,保证算法的准确性、实时性,同时实现目标的检测和分割有重要的研究意义。

本文基于深度学习的目标检测算法和图像分割算法,用 ResNet-50 作为共享的特征提取网络,设计了能同时进行目标检测与分割的多任务联合方法。从检测和分割的准确性、实时性,时间的同步性等方面考虑,在联合方法中使用了改进的 SSD 目标检测算法和基于 DeepLab v3 的路面分割算法。其中分割方案只对路面进行分割识别得到车辆的基础可行驶区域,在提高分割计算效率的同时,对检测算法判别物体的位置信息进行辅助判断。用自建的检测和分割混合数据集对本方法进行验证实验及实车验证,结果表明,本方法能在保证高准确性和实时性的情况下,完成目标的检测和分割,并给出车辆的基础可行驶区域。

## 2 多任务联合方法模型结构及实现过程

### 2.1 模型结构

#### 1) 特征提取网络

不同特征提取网络检测的准确率和实时性都有较大差异,因此选择合适的编码器即公用的特征提取网络尤为重要,从 AlexNet 的 8 层网络到 VGG 的 16 层网络,不同深度的卷积网络整合了不同层次的特征。Liu 等<sup>[8]</sup>将 VGG 网络直接作用于特征图,以预测多目标类别和外围框,同时采用了 SSD 算法,提高了检测的实时性和准确性,但该方法对图像中的小物体和重叠物体的识别效果较差。因此人们

通过加深网络的深度获取更细微的特征,以提高对小物体的识别精度。Simonyan 等<sup>[9]</sup>证明了网络层数在加深的同时会使反向传播失败,导致网络训练不能顺利进行,虽然可通过归一化<sup>[10]</sup>和 dropout<sup>[11]</sup>技术解决这些问题,但还是会出现检测精度退化或检测精度饱和的现象。

为解决网络加深导致的精度退化问题,He 等<sup>[12]</sup>提出了深度残差网络(ResNet)模型。如图 1 所示,在 ResNet 中,导入了“快捷结构”,该结构跳过一个或多个卷积层将自身映射  $x$  与残差函数由  $F(x)$  相互叠加,再经过一个非线性的 ReLU 激活函数,组成完整的残差学习模块。该模块可无衰减地传递反向传播信号,即使加深网络层数,也能进行高效的学习。

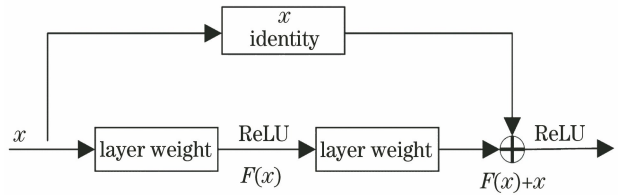


图 1 ResNet 的结构

Fig. 1 Structure of ResNet

由于 ResNet-50 中的参数量较少,避免了网络因参数过多出现的过拟合现象,且在增加网络深度的同时不会发生网络退化,提高了模型的泛化能力。该网络可对图像进行下采样,产生多尺度的特征信息,并将其传输给检测和分割编码器进行图像的检测和分割。因此使用 ResNet-50 作为特征编码器中的特征提取网络,其网络结构如表 1 所示,其中 Conv 表示卷积核。

表 1 ResNet-50 结构图

Table 1 Structure diagram of ResNet-50 network

Layer	Output size	Conv layer
Pre-block	$160 \times 120$	$[7 \times 7, 64]$
Block1	$80 \times 60$	$\begin{bmatrix} 1 \times 1, & 64 \\ 3 \times 3, & 64 \\ 1 \times 1, & 256 \end{bmatrix} \times 3 \text{ Conv}$
Block2	$40 \times 30$	$\begin{bmatrix} 1 \times 1, & 128 \\ 3 \times 3, & 128 \\ 1 \times 1, & 512 \end{bmatrix} \times 4 \text{ Conv}$
Block3	$20 \times 15$	$\begin{bmatrix} 1 \times 1, & 256 \\ 3 \times 3, & 256 \\ 1 \times 1, & 1024 \end{bmatrix} \times 6 \text{ Conv}$
Block4	$10 \times 8$	$\begin{bmatrix} 1 \times 1, & 512 \\ 3 \times 3, & 512 \\ 1 \times 1, & 2048 \end{bmatrix} \times 3 \text{ Conv}$

2) 模型结构

MultiNet 模型<sup>[13]</sup>效率较高,且具有通用性,便于对网络结构进行改进,因此实验基于 MultiNet 模型设计了多任务联合检测和分割方法结构。其中检

测和分割算法共同使用一个编码器(特征提取网络),两个分支任务解码器(检测和分割算法)使用编码器中提取的多尺度特征分别进行检测和分割任务。总体框架如图 2 所示。

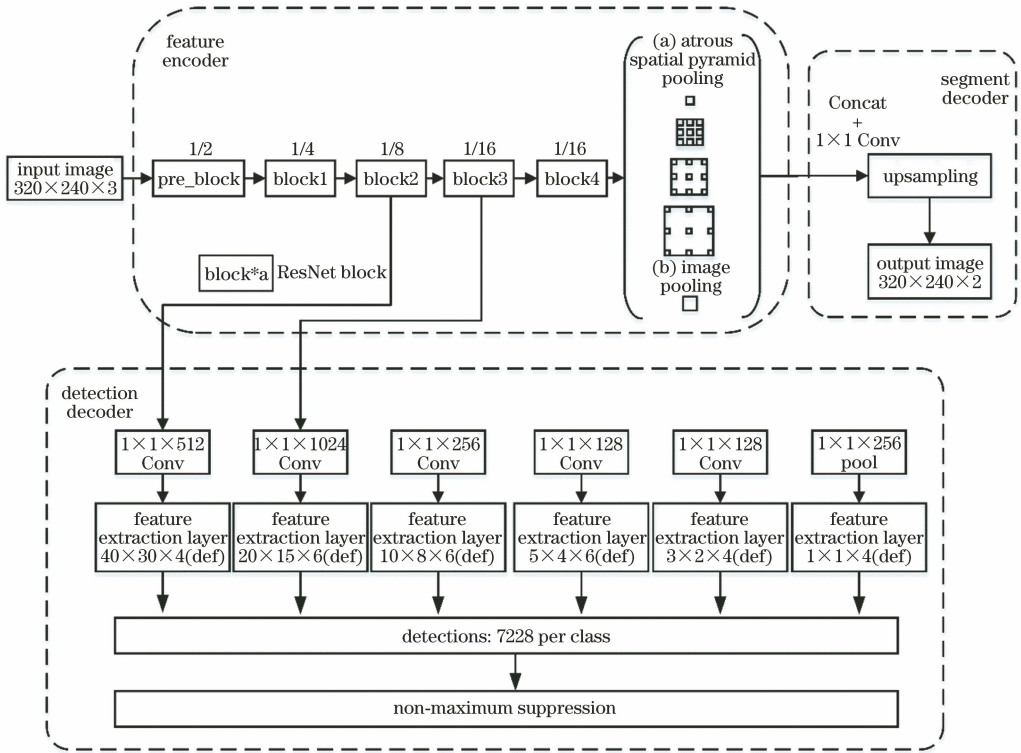


图 2 图像检测和分割联合方法网络结构

Fig. 2 Network structure of joint method for image detection and segmentation

由于车辆在行驶中拍摄的图像各类目标大小不一致,所以在相同尺度下,检测算法对各类目标的检测精度会有差异,大目标需要使用小尺度才能获得更好的特征信息,小目标需要使用大尺度才能获得更好的特征信息,因此对 SSD 算法进行了改进,设计成编码-解码结构。编码器采用 ResNet-50 作为特征提取网络,通过逐渐减小特征图的空间尺寸,在更深的编码器输出中捕获更多的信息。解码器采用 SSD 算法的多尺度特征预测和快速回归思想,对尺度大小不同的特征图直接用卷积提取检测结果。图 2 中的 feature encoder 单元和 detection decoder 单元,构成了检测算法的网络结构。

此外,车辆拍摄的图像易受环境影响,导致路面在同一张图像中的不同位置有不同的颜色、光度等特征。为了在分割过程中获取更多的特征信息,提高分割准确率,模型将并行多孔空间金字塔池化(ASPP)与 ResNet-50 相结合,改进了传统的 DeepLab v3 分割算法。其中 ASPP 对图像特征进行多尺度映射,以获取更加丰富的上下文信息,从而提高模型对路面基础

可行驶区域的分割效果。由于只需要对路面和背景进行分割,所以输出图像设定为 2 通道。

2.2 实现过程

1) 检测算法的实现

为便于计算,将图像尺寸从 1288 pixel × 964 pixel 降为 320 pixel × 240 pixel,如图 3 所示。可以看出 SSD 算法采用多个尺度大小不同的特征图进行检测,如图 3(b)中尺度为 10 × 8 的特征图,每个单元的先验尺度(框①)比图 3(a)中尺度为 20 × 15 特征图中每个单元的先验尺度(框②)更接近于真实目标(框③),所以大尺度特征图用来检测小目标,而小尺度特征图用来检测大目标。

将多尺度图像分成 M × N 大小的单元,如图 4 所示。特征图中各单元的任意一个先验框,都会输出一套独立的检测值,对应一个边界框,主要分为两部分。第一部分是各个类别的置信度值,假设检测目标的类别个数为 c,由于检测过程中背景也作为一类存在,所以最终需要预测 c + 1 个置信度值。下文提及的 c 个类别置信度值中包含了背景类别,

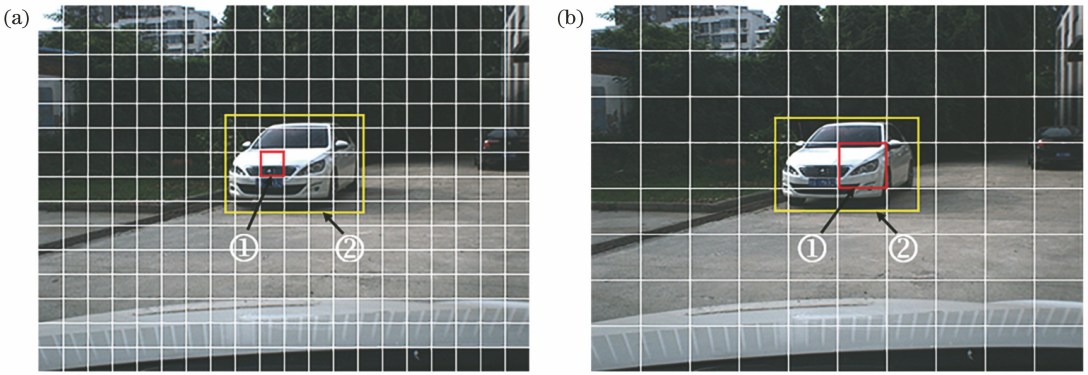


图 3 不同尺度的特征预测图;(a)  $20 \times 15$ ; (b)  $10 \times 8$

Fig. 3 Feature prediction maps of different scales; (a)  $20 \times 15$ ; (b)  $10 \times 8$

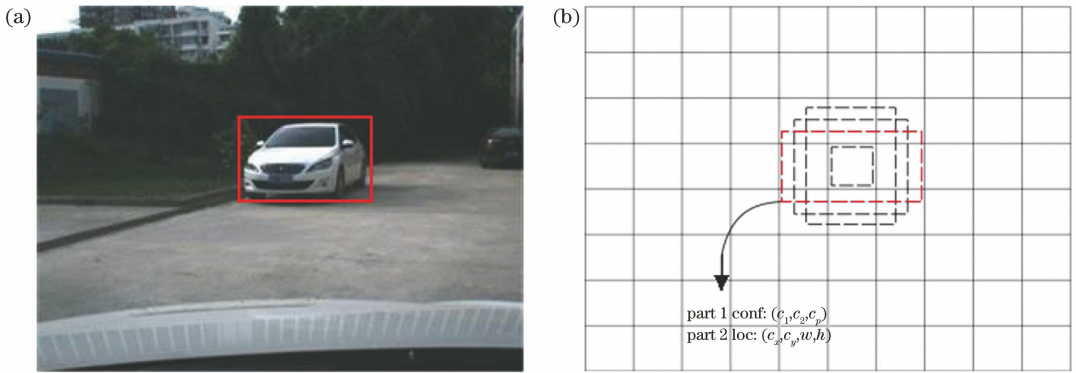


图 4 SSD 的先验框。(a)带真实目标(GT)框的图像;(b)尺寸为  $10 \times 8$  的特征图

Fig. 4 Prior box for SSD. (a) Image with ground truth box; (b) feature image with a size of  $10 \times 8$

即真实的检测类别只有  $c - 1$  个。第二部分是边界框 loc, 包含的  $(c_x, c_y, w, h)$  四个值, 其中  $c_x, c_y, w, h$  分别表示检测框的中心坐标以及宽高。假设每个单元设置的先验框数目为  $K$ , 则每个单元需要预测  $c$  个类别置信度值和 4 个坐标值  $(c_x, c_y, w, h)$ , 即  $(c + 4)K$  个预测值, 对于尺寸为  $M \times N$  的特征图, 所有的单元共需要  $(c + 4)KMN$  个预测值, 由于 SSD 用卷积做检测, 所以就需要  $(c + 4)K$  个卷积核完成该特征图的检测, 整个检测流程如图 5 所示的 detection 部分, 其中 NMS 表示非极大值抑制。

### 2) 分割算法实现过程

ResNet-50 将输入图像下采样为原图的  $1/16$  后, 输入到 ASPP 中进行多尺度采样。ASPP 中的带孔卷积采用不同的步长将输入图像直接划分为不同的尺寸, 提取不同尺度的特征。同时增加了分支对最后一个残差块输出的特征图进行全局池化, 获取一个图像级别的特征图以捕获上下文信息。将所有尺度特征融合后, 通过双线性上采样和批次归一化处理获得与原始图像尺寸大小相同的分割结果。整个分割流程如图 5 中的 segmentation 部分。

### 2.3 损失函数

在训练中损失函数是评价神经网络性能“恶劣程度”的指标, 检测算法的总体目标损失函数可表示为位置损失  $L_{loc}$  与置信度损失  $L_{conf}$  的加权和。

$$L_{det}(x, c, l, g) = \frac{1}{W} [L_{conf}(x, c) + \alpha L_{loc}(x, l, g)], \quad (1)$$

式中,  $W$  为先验框的正样本数量,  $c$  为类别置信度的预测值,  $\alpha$  为权重系数, 设置为 1,  $l$  为先验框所对应边界框的位置预测值,  $g$  为真实标签的位置参数。位置损失可表示为

$$L_{loc}(x, l, g) = \sum_{i \in X_{Pos}} \sum_m x_{ij}^l X_{smoothL1}(l_i^m \hat{g}_j^m), \quad (2)$$

其中

$$X_{SmoothL1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{others} \end{cases}, \quad (3)$$

式中,  $\hat{g}_j^m$  为真实标签的近似回归预测框,  $l_i^m$  为先验框的位置预测值, 其中  $m$  为  $c_x, c_y, w, h, x_{ij}^l \in \{1, 0\}$  为一个指示参数, 当  $x_{ij}^l = 1$  时表示第  $i$  个先验框与第  $j$  个真实标签匹配, 且真实标签的类别为  $p$ , SSD 的预测框和真实标签框之间的偏移回归类似于

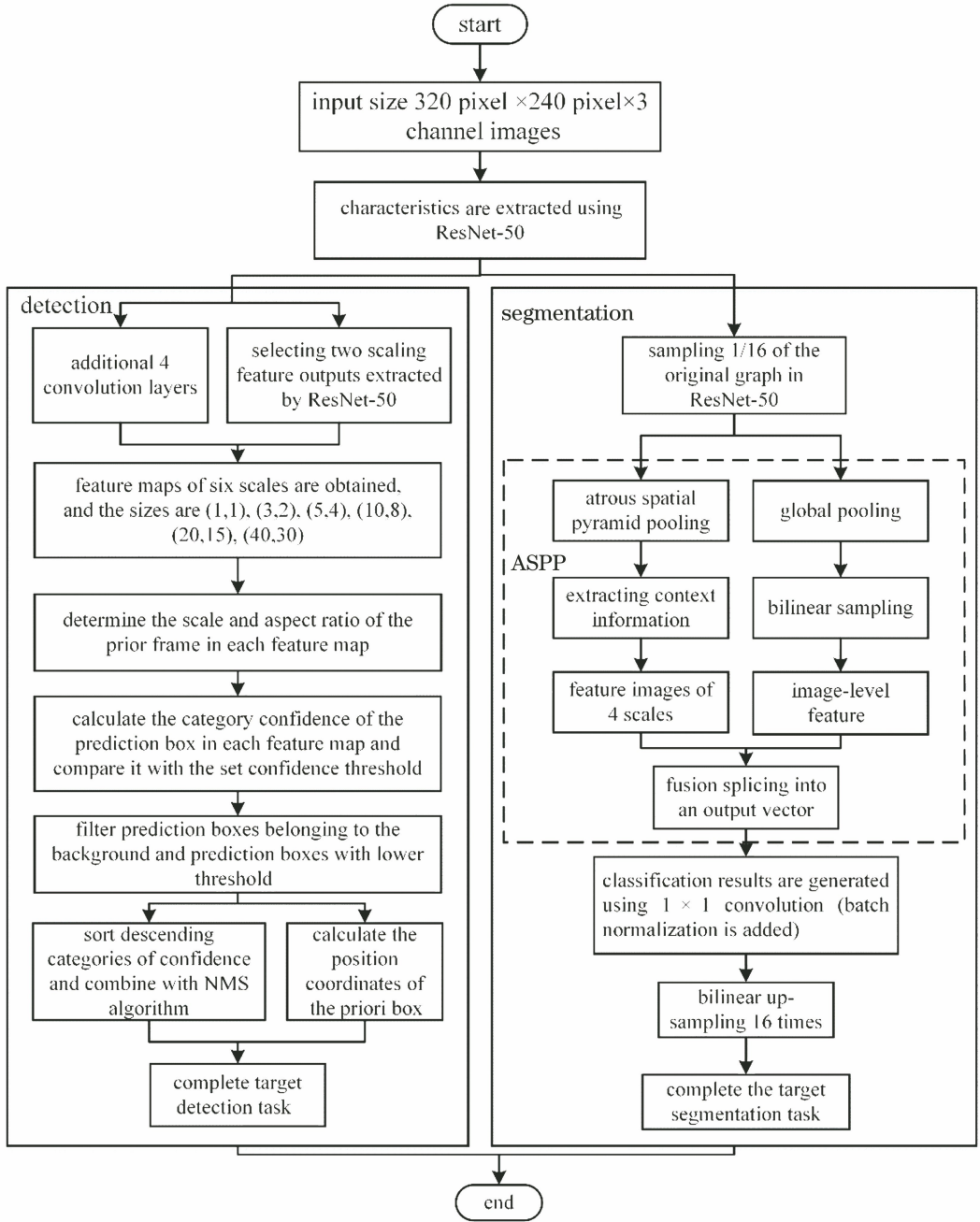


图 5 多任务联合方法流程框图

Fig. 5 Flow chart of multi-task joint method

Faster R-CNN<sup>[14]</sup>中的 anchor 理念,可用平移向量  $d_{i^x}^c$ 、 $d_{i^y}^c$  和尺度缩放因子  $d_i^w$ 、 $d_i^h$  获取

$$\begin{aligned} \hat{g}_{j^x}^c &= (g_{j^x}^c - d_{i^x}^c) / d_i^w & \hat{g}_{j^y}^c &= (g_{j^y}^c - d_{i^y}^c) / d_i^h \\ \hat{g}_j^w &= \log(g_j^w / d_i^w) & \hat{g}_j^h &= \log(g_j^h / d_i^h), \end{aligned} \quad (4)$$

置信损失可表示为

$$L_{\text{conf}}(x, c) = - \sum_{i \in X_{\text{Pos}}} x_{ij}^p \log(\hat{c}_i^p) - \sum_{i \in X_{\text{Neg}}} \log(\hat{c}_i^0), \quad (5)$$

式中,  $\hat{c}_i^p = \exp(c_i^p) / \sum_p \exp(c_i^p)$ ,  $c_i^p$  为类别  $p$  的第  $i$

个预测框置信度,  $X_{\text{Pos}}$  为正样本,  $X_{\text{Neg}}$  为负样本。

语义分割提取的实质是对图像中的像素点进行分割,采用交叉熵损失函数对图像中的像素点进行分割计算,可表示为

$$L_{\text{seg}} = - \frac{1}{|A|} \sum_{a \in A} \sum_{b \in B} q_a(b) \log[p_a(b)], \quad (6)$$

式中,  $L_{\text{seg}}$  为基础可行驶区域分割的损失函数,  $B$  为需要分割的所有类的集合,  $A$  为某批次中的一份样本,  $p_a(b)$  为网络预测的像素点类别,  $q_a(b)$  为真实标记的像素点类别。

联合算法的损失函数为两个任务的损失函数和,可表示为

$$L_{\text{multi}} = L_{\text{det}} + L_{\text{seg}}, \quad (7)$$

式中,  $L_{\text{det}}$  为目标检测的损失函数,  $L_{\text{seg}}$  为目标分割的损失函数。

## 2.4 评价指标

实验采用的评价标准为常用的公开标准,混合算法中的检测算法和分割算法分别选用平均精度均值(mAP)、平均交并比(MIoU)和每秒传输帧数(FPS)进行评价。

### 1) 平均精度均值

查准率  $P(\%)$  和召回率  $R(\%)$  分别表示预测的检测框中检测正确的比例和预测框中正确框在所标注框中的比例,可以表示为

$$P = \frac{X_{\text{TP}}}{X_{\text{TP}} + X_{\text{FP}}} \times 100\%, \quad (8)$$

$$R = \frac{X_{\text{TP}}}{X_{\text{TP}} + X_{\text{FN}}} \times 100\%, \quad (9)$$

式中,  $X_{\text{TP}}$  为各个类别检测到的正确检测框的数量,  $X_{\text{FP}}$  为检测到的错误检测框数量,  $X_{\text{FN}}$  为未检测到错误检测框数量。平均精度(AP)是以召回率  $R$  为横轴、查准率  $P$  为纵轴绘制的  $P$ - $R$  曲线与横纵轴所围成的面积, mAP 是测量网络对目标物体敏感度的一种标准, mAP 越高说明识别精度越高,因此实验选用 mAP 为检测算法的评价标准。

### 2) 平均交并比

MIoU 是计算网络预测分割与真实分割的交并比,可表示为

$$X_{\text{MIoU}} = \frac{1}{z+1} \sum_{i=0}^z \frac{p_{ii}}{\sum_{j=0}^z p_{ij} + \sum_{j=0}^z p_{ji} - p_{ii}}, \quad (10)$$

式中,  $p_{ii}$  为分类正确的正例数量,  $z$  为除了背景类别外还有  $z$  类物体,  $p_{ij}$  为原本属于第  $i$  类却被识别为第  $j$  类的像素数量。MIoU 是最具有代表性的语义分割评价指标,因此,采用 MIoU 作为路面基础可行驶区域分割的评价指标。

### 3) 每秒传输帧数

FPS 表示每秒能处理图像的帧数,是评判一个网络进行目标检测的速度指标。FPS 越大,表明网络的检测速度越快。

## 3 模型训练及仿真实验结果分析

### 3.1 仿真环境

基于 TensorFlow 深度学习框架对模型进行训

练,实验环境:操作系统为 Linux 下的 Ubuntu 16.04、TensorFlow 深度学习框架,CPU 为 Intel i7-7700k,GPU 为 NVIDIA GTX 1080Ti,固态硬盘容量为 1 T,CUDA 8.0 版本并行计算框架配合深度神经网络加速库、Python 3.5、运行内存为 64 G。

### 3.2 混合数据集

多任务联合方法的输入图像为检测与分割的融合数据集,包含了公开 Cityscapes 数据集、CamVid 数据集中的部分数据以及自采集数据。由于两个数据集使用的基础图像一致,只是增加了不同类型的标注。因此,联合方法在 TensorFlow 中用队列直接读取两种类型的数据集,然后以分批的形式进行训练。该混合数据集包括 6000 张尺寸为 1288 pixel×964 pixel 的图像,其中训练集有 4000 张,交叉验证集有 1000 张,测试集有 1000 张。

### 3.3 数据处理及关键参数确定

#### 1) 数据增强和数据的随机批处理

为提高模型的泛化能力,对采集的行驶环境图像使用水平翻转、随机裁剪和色彩抖动等方法<sup>[15]</sup>进行数据增强。同时在训练过程中,使用随机批处理训练机制,防止发生过拟合现象,以提高网络模型的收敛速度。

#### 2) 关键训练参数的确定

在训练开始时设置一个较小的学习率,每次更新网络的同时增加学习率,描绘学习率与损失值的变化曲线<sup>[16]</sup>。通过对比分析,将初始学习率设置为 0.01,用 Adam 法<sup>[17]</sup>对学习率进行优化衰减。网络中的权重衰减系数为  $5 \times 10^{-4}$ ,解码器内部卷积的 dropout 为 0.5。

### 3.4 仿真实验结果分析

#### 1) 仿真结果分析

实验在 TensorFlow 网络框架下总共进行了 50000 步训练,训练步数为横坐标,学习率和训练损失分别作为纵坐标,拟合得到的训练过程曲线如图 6 所示。

从图 6 中可以看出,在前 10000 步中衰减比较迅速,之后逐渐平缓,整个训练过程的衰减较为平整。而 loss 值在网络训练初期特别大,训练过程中大幅度降低,最终趋于稳定。同时,训练中没有出现梯度爆炸、梯度弥散等问题,且在前 10000 步下降速度很快,这表明该网络具有快速收敛特性。随着学习率的逐渐衰退,损失函数的下降速度变慢,训练完成后得到一个较好的权重,将该权重放入联合方法中,对 1000 张测试图像进行测试,联合方法对测试图像的检测分割结果如图 7 所示。

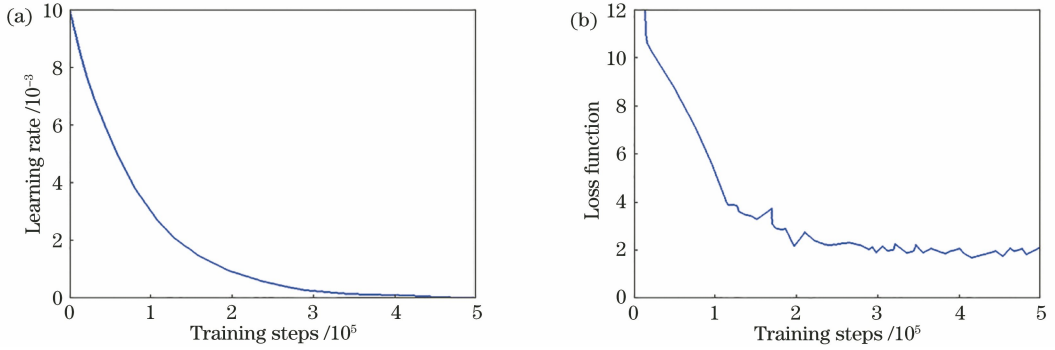


图 6 学习率和损失函数曲线图。(a)学习率;(b)损失函数

Fig. 6 Graph of learning rate and loss function. (a) Learning rate; (b) loss function

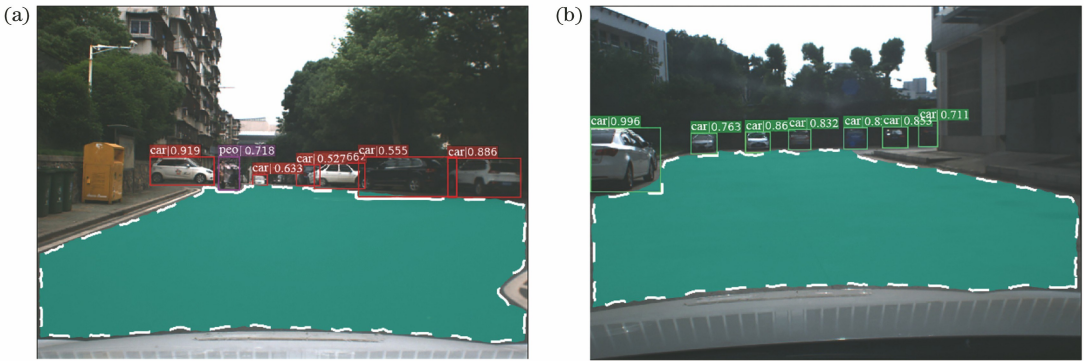


图 7 测试数据集在联合方法中的检测和分割效果图。(a)场景 A;(b)场景 B

Fig. 7 Detection and segmentation of data in the joint method. (a) Scene A; (b) scene B

联合方法采用矩形检验框来检测目标物体,采用虚线围出道路的基础可行使区域。从图 7 可以看出,联合方法能准确地框选出目标物体,也能够准确分割

出道路的基础可行使区域,但分割的边缘还存在一些误差,但总体上联合方法的检测和分割效果较好,不同方法的检测性和分割性分别如表 2 和表 3 所示。

表 2 不同方法的检测性

Table 2 Detection ability of different methods

Detection model	Feature network	mAP / %	FPS /frame	Dataset
Faster R-CNN	VGG16	89.23	7	cityscapes
Faster R-CNN	ResNet-101	91.80	8	cityscapes
YOLO v2	GooleNet	83.46	21	cityscapes
Ours	ResNet-50	90.15	32	mixed dataset

表 3 不同方法的分割性

Table 3 Segmentation ability of different methods

Split model	MIoU	FPS /frame	Dataset
PSPNet	76.6	2	cityscapes
ResNet-38	86.2	5	cityscapes
DeepLab v3	77.8	8	cityscapes
Ours	87.6	32	mixed dataset

从表 2 中看出,本方法对混合测试数据集的检测的 mAP 比 YOLO v2 检测算法高,比基于 ResNet-101 的 Faster R-CNN 检测算法低 1.65 个百分点;从表 3 中看出,相比其他分割方法,本联合方法的分割 MIoU 较高,表现出较好的分割效果,FPS 达到 32 frame,基本实现了实时检测,这表明本方法有良

好的实时性。

### 2) 实车实验验证

为测试该方法的实用效果,将联合方法最终训练完成的模型移植到智能车车载工控机中,在行驶中测试该方法的性能。输出图像中用矩形方框识别道路上的目标,并得到其分类结果。用虚线围出道路的基础可行使区域,结果如图 8 所示。可以看出,本方法可以完成目标的正确分类,且分割算法正确地对道路进行了分割,并给出了基础的可行使区域;联合方法运行的过程中,检测 mAP 为 89.00%,分割 MIoU 为 83.0,FPS 稳定在 31 frame。这表明本方法具备实时检测与分割的能力,可在动态环境中

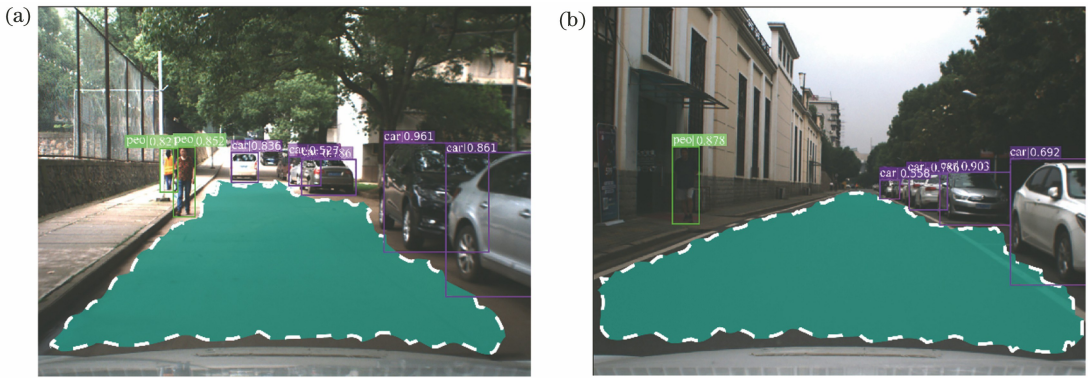


图 8 智能车联合方法实际效果图。(a)场景 A;(b)场景 B

Fig. 8 Actual effect figure of intelligent vehicle joint method. (a) Scene A; (b) scene B

对目标物进行检测与分割。

## 4 结 论

由于无人驾驶车辆在运行过程中需要搭载多任务系统,并行任务使有限的计算资源承受了巨大的压力,从而导致视觉感知过程中存在滞后、准确率低等问题。为保证多任务算法在时间上同步的同时减少网络参数运算量,且检测和分割的准确性、实时性满足要求,提出了一种多任务的联合检测与分割方法,设计了编码-解码的网络模型,选用网络中参数量较少的 ResNet-50 作为检测和分割的共用编码器,避免了训练过程中因参数过多出现的过拟合和随着深度的增加网络退化的问题。检测解码器采用 SSD 的多尺度特征图预测和快速回归思想,设置了多个先验框确保实时性和准确率。分割解码器采用 DeepLab v3 中的 ASPP,对图像进行多尺度采样,经全局池化、特征融合得到分割结果。实验结果表明,相比单独的目标检测和分割方法,本方法能在检测和分割的过程中,保持较高的识别准确性和实时性,同时减少了算法在训练过程中的参数数量,满足智能车的相关任务要求。

## 参 考 文 献

- [1] Li S S. The research of multi-object detection in traffic scene based on deep learning [D]. Changsha: Hunan University, 2017.  
李珊珊. 基于深度学习的交通场景多目标检测 [D]. 长沙: 湖南大学, 2017.
- [2] Zhao H, An W S. Image salient object detection combined with deep learning [J]. Laser & Optoelectronics Progress, 2018, 55(12): 121003.  
赵恒, 安维胜. 结合深度学习的图像显著目标检测 [J]. 激光与光电子学进展, 2018, 55(12): 121003.

- [3] Xu L X, Chen X J, Ban Y, et al. Method for intelligent detection of parking spaces based on deep learning [J]. Chinese Journal of Lasers, 2019, 46(4): 0404013.  
徐乐先, 陈西江, 班亚, 等. 基于深度学习的车位智能检测方法 [J]. 中国激光, 2019, 46(4): 0404013.
- [4] Zhang X F, Liu J, Shi Z S, et al. Review of deep learning-based semantic segmentation [J]. Laser & Optoelectronics Progress, 2019, 56(15): 150003.  
张祥甫, 刘健, 石章松, 等. 基于深度学习的语义分割问题研究综述 [J]. 激光与光电子学进展, 2019, 56(15): 150003.
- [5] Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(4): 640-651.
- [6] Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [7] Chen L C, Papandreou G, Kokkinos I, et al. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector [M] // Leibe B, Matas J, Sebe N, et al. Computer Vision-ECCV 2016. Lecture Notes in Computer Science, Cham: Springer, 2016, 9905: 21-37.
- [9] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [EB/OL]. [2019-08-30]. <https://arxiv.org/abs/1409.1556>.



- [10] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[EB/OL]. [2019-08-28]. <https://arxiv.org/abs/1502.03167>.
- [11] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [12] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [13] Teichmann M, Weber M, Zollner M, et al. MultiNet: real-time joint semantic reasoning for autonomous driving [C] // 2018 IEEE Intelligent Vehicles Symposium, June 26-30, 2018, Changshu, Jiangsu, China. New York: IEEE, 2018: 1013-1020.
- [14] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C] // Proceedings of the 25th International Conference on Neural Information Processing Systems-Volume 1, December 3-6, 2012, Red Hook, NY, US. New York: ACM, 2012: 1097-1105.
- [16] Smith L N. Cyclical learning rates for training neural networks [C] // 2017 IEEE Winter Conference on Applications of Computer Vision, March 24-31, 2017, Santa Rosa, CA, USA. New York: IEEE, 2017: 464-472.
- [17] Kingma D P, Ba J. Adam: a method for stochastic optimization[EB/OL]. [2019-09-01] <https://arxiv.org/abs/1412.6980>.