文章编号：2095-2163(2020)03-0095-04 中图分类号：TP391.41 文献标志码：A

基于 Faster RCNN 的行人及车辆类型检测

邵丽萍[1]，魏相站[1]，李春红[1]，唐志英[2]，白忠臣[2]，张正平[1]
（1 贵州大学 大数据与信息工程学院，贵阳 550025；2 贵州大学 贵州省光电子技术及应用重点实验室，贵阳 550025）

摘 要：随着汽车数量与日俱增，交通事故的发生频次也在增加，针对车辆类型和行人的检测问题，本文在原始 Faster RCNN 的基础上，首先使用残差网络 RES101 代替传统的 VGG16 网络作为共享卷积层，进行图像特征的提取，然后改变原来的锚框尺寸方案，使用锚框尺寸为 4、8、16 代替原来锚框尺寸，得到行人及车辆类型检测模型。通过在 KITTI 测试集上的测试结果表明，使用本文模型平均检测准确率可达 86.5%，相比原始 Faster RCNN 平均准确率提高了 3.65%，相比于使用残差网络 RES101 作为卷积层的 Faster RCNN 平均准确率提高了 2.06%。

关键词：Faster RCNN；残差网络；特征提取；锚框选区

Pedestrian and vehicle type detection based on Faster RCNN

SHAO Liping[1], WEI Xiangzhan[1], LI Chunhong[1], TANG Zhiying[2], BAI Zhongchen[2], ZHANG Zhengping[1]
(1 College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;
2 The Key Laboratory for Photoelectric Technology and Application, Guizhou University, Guiyang 550025, China)

【Abstract】 With the increasing number of vehicles and frequent traffic accidents, aiming at the detection of pedestrians and vehicle types, based on the original Faster RCNN, firstly, the residual network RES101 is used instead of the traditional VGG16 network as the shared convolutional layer to extract the image features. Then, the original anchor frame size scheme is changed, and the size of the anchor frame is 4, 8 and 16 instead of the size of the original anchor frame, and the pedestrian and vehicle type detection model are obtained. The test results on the KITTI test set show that the average detection accuracy of the proposed model is 86.5%, 3.65% higher than the original Faster RCNN, and 2.06% higher than the original Faster RCNN using residual network RES101 as the convolutional layer.

【Key words】 Faster RCNN; residual network; feature extraction; anchor frame selection

0 引 言

目前，随着城市汽车数量的增多，道路交通流量在不断地增加，对汽车驾驶的安全性也提出了更高的要求。而自动驾驶作为汽车辅助驾驶的系统，能够确保汽车在行驶途中的安全，现已成为当下的热门实用研发技术之一。自动驾驶的主要技术分为行人检测、碰撞检测及夜视辅助等，而碰撞检测与行人检测的实现依赖于计算机视觉技术中的图像识别技术对汽车行驶途中的车辆和行人进行识别。在汽车行驶过程中，计算机只能识别到目标图像的 RGB 像素矩阵，为了得到较好的识别效果，本次研究中使用了 Faster RCNN[1-2] 算法对汽车行驶路线中的车辆和行人进行识别[2-4]。基于此，文中提出了一种改进 Faster RCNN 的目标检测方法，使用 ResNet-101 深度残差网络代替传统的 VGG16 网络作为特征提取网络，并且调整锚框尺寸大小来提高检测的准确率，最后在 KITTI 数据集上进行测试。

1 Faster RCNN 模型简介

1.1 Faster RCNN 结构

为了使检测算法能够对车辆类型进行快速有效地定位和检测，使用 ResNet-101 代替传统的 VGG16 作为共享卷积层，并且对区域建议网络中最终生成的感兴趣区域数量进行调整，使算法在保证准确率的基础上进一步提高检测速度。模型整体结构如图 1 所示。由图 1 可知，该模型结构中的各主要部分的功能描述具体如下：对数据集图像进行特

基金项目：国家自然科学基金(61865002)；贵州省科技支撑计划(SY[2017]2881)；贵州大学引进人才项目(201602)；贵州省人才团队项目([2018]5616)；中央引导地方科技发展专项(QKZYD[2017]4004)。

作者简介：邵丽萍(1996-)，女，硕士研究生，主要研究方向：通信与信息系统；魏相站(1995-)，男，硕士研究生，主要研究方向：图像处理；李春红(1994-)，女，硕士研究生，主要研究方向：通信与信息系统；唐志英(1995-)，女，硕士研究生，主要研究方向：医学信息工程；白忠臣(1979-)，男，博士，副教授，主要研究方向：纳米传感器；张正平(1964-)，男，博士，教授，主要研究方向：电磁场与微波技术。

通讯作者：张正平 Email：zpzhang@gzu.edu.cn

收稿日期：2019-12-30

征提取生成特征图;使用 RPN 区域来调整候选框并得到调整好的候选框;通过 ROI 池化层得到固定大小的兴趣区域;送入全连接层和 softmax[5] 计算求得每个候选框的所属类别,输出类别的得分;同时再次利用框回归获得每个候选区相对实际位置的偏移量预测值,用于对候选框进行修正,得到更精确的目标检测框。

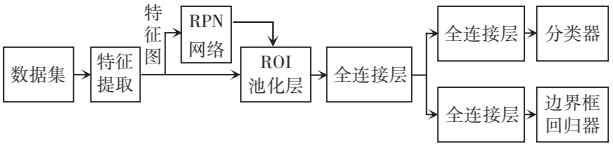

图 1 Faster RCNN 模型结构图

Fig. 1 Faster RCNN model structure diagram

1.2 RPN 网络

RPN 网络是 Faster R-CNN 的核心,是一种全卷积网络,用来对目标检测网络进行选择性搜索。经过卷积神经网络进行特征提取的特征图输入到 RPN 网络,先进行一次 3 * 3 的卷积运算,再分别进行 2 次 1 * 1 的卷积运算。其中,一个是计算检测区域是前景或背景的概率,用来给 softmax 层进行前景或背景分类;另一个是用于给候选区域精确定位[6]。RPN 网络使用滑动窗口,可同时预测多个候选区,滑动每一个滑动窗口后都会产生一个特征向量,将产生的特征向量传送到全连接层即可判断检测目标的位置和类别[7]。RPN 网络的结构如图 2 所示,RPN 网络共有 K 个锚框、K 个区域建议框、$2K$ 个对应分类层输出及 $4K$ 个对应回归层输出,其中的分类层输出用来指示非目标与目标的概率,回归层输出用来标注区域建议框的位置。

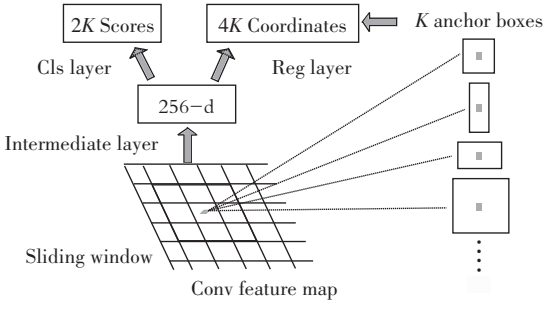

图 2 RPN 网络结构图

Fig. 2 Structure of RPN Network

2 改进部分

2.1 锚框尺寸调整

由于车辆类型和人的检测受建筑物和树木遮挡等的影响,导致车辆和人的显示尺寸差异比较大,其长宽比也是复杂多样。原始的 Faster RCNN 模型包含 9 种锚框,其长宽比分别为 0.5,1,2,尺寸分别为 8,16,32。如果按照原始 Faster RCNN 模型的锚框方案,输入图像在经过池化后,特征图中的各点对应的感受野尺寸为 16×16。使用最小尺度映射的锚框尺寸都达到 128,而实际中存在一些距离较远的行人以及车辆,进而其占有的尺寸也比较小,由于这些较小尺寸的目标在测试时可能会出现一定的定位偏差,进而造成检测错误,准确率也会随之降低。因此,根据人和车辆尺寸差异比较大、长宽比更加复杂多样的特点,即可调整原始 Faster RCNN 模型中的 RPN 网络的锚框尺寸。调整后锚框的种类保持不变,只是将锚框尺寸改成 4,8,16,有助于增强对距离远的行人以及车辆的检测。测试结果表明,经过调整后的锚框尺寸可以使得检测准确率提高。锚框尺寸对比见表 1。

表 1 锚框尺寸选择方案

Tab. 1 Selection scheme of anchor frame size

锚框尺寸	选择方案
原始尺寸	[8,16,32]
调整尺寸	[4,8,16]

2.2 ResNet 网络

在深层网络提取的特征图中,远距离的检测目标特征提取量很少,这就需要对特征提取网络做出改进,让改进后的特征提取网络获取图像中更多的小尺寸物体的特征。由于残差网络[8]在网络卷积中加入大量的跳跃连接,使其能够在训练较深的网络中提取更多小尺寸物体特征。ResNet 网络已经成功训练出了 152 层神经网络,还加入了残差模块(Residual block),在网络深度增加的同时,有效地保证了模型的准确度。ResNet 网络结构如图 3 所示。由图 3 可知,在网络结构中添加了直连通道来实现隔层连接,将初始的输入数据传递到后面的网络层中。对于输入的 x 期望,假设所要求的映射关系为 $H(x)$,在求解过程中,$H(x)$ 比 $F(x)$ 复杂得多,因而研究可以通过求出 $H(x)$ 的残差形式,也即 $H(x)=F(x)-x$ 来实现。

研究推得残差结构块的数学公式可写为:

$$y_1 = h(x_1) + F(x_1, W_1), \tag{1}$$

$$x_{l+1} = f(y_l), \tag{2}$$

其中,x_l 表示第 l 个残差单元的输入;x_{l+1} 表示第 l 个残差单元的输出;F 表示残差函数;f 表示 $ReLU$ 激活函数;恒等映射即为 $h(x_l)=x_l$,则从 l 到 L 的学习特征为:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, W_i), \tag{3}$$

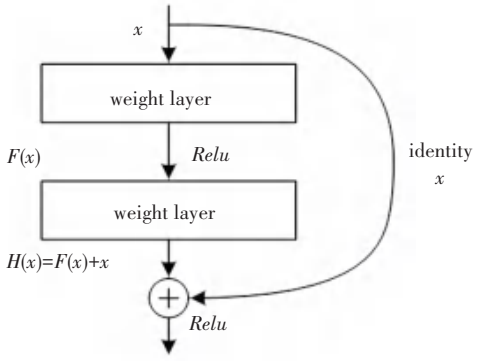

图 3　ResNet 网络结构图

Fig. 3　ResNet network structure diagram

因此，反向传播梯度可用下式求出：

$$\frac{\partial loss}{\partial x_l}=\frac{\partial loss}{\partial x_L}\times\frac{\partial x_L}{\partial x_l}=\frac{\partial loss}{\partial x_L}\times\left(1+\frac{\partial}{\partial x_L}\sum_{i=l}^{L-1}F(x_i,W_i)\right).\tag{4}$$

综上分析可知，应用此方法可以有效地对梯度进行无损传播，避免了梯度的消失问题，因而在上述模型中对汽车行驶途中的车辆类型及行人的图像具有良好的特征提取效果。

3　实验及结果分析

本文算法采用 KITTI 数据集。KITTI 数据集是由德国卡尔斯鲁厄理工学院和丰田美国技术研究院共同创建，是世界上最大的自动驾驶仪场景中适合自动驾驶计算机视觉算法的数据集[9]。本文选用 2D 的数据集，KITTI 数据集包含市区、乡村和高速公路等场景采集的真实图像数据，图像数据的平均分辨率为 1 240 * 375，每张图像中最多达 15 辆车和 30 个行人，还有各种程度的遮挡与截断，整个数据集分为 8 个类别：car，van，truck，pedestrian，pedestrian(sitting)，cyclist，tram 以及 misc。本文将原数据集中的 pedestrian，pedestrian(sitting)归为 pedestrian 一类，去除 misc、cyclist 的目标类别，同时将目标分为 car，van，truck，pedestrian，tram 五类，并在此数据集上验证本文算法的有效性。该数据集一共有 7 480 张图片，按 8:1:1 分成训练集、验证集以及测试集。本文实验算法基于 PyTorch 深度学习框架[10]实现，在搭载 NVIDIA Quadro P5000 GPU 的 Ubuntu 16 系统的实验配置下完成。训练网络的初始学习率为 0.001，模型训练的批量为 1，衰减因子为 0.8，在训练过程中，模型的损失值会伴随着迭代次数的增加而减少，直到模型的损失值趋于稳定、甚至更优的状态。在本文中总迭代次数为 150 000 次，模型损失值曲线如图 4 所示，此时的模型损失曲线几乎处于稳定状态，表明模型已经收敛。训练好的模型检测结果示例如图 5 所示，该模型分别将图片中的行人以及车辆检测出来，并给出预测类别以及预测类别的概率。

通过训练好的模型对测试集进行测试，测试结果见表 2。结果显示，使用传统的 Faster RCNN 网络模型的平均检测准确率为 82.85%，平均每张图像的检测时间为 0.200 2 s；改用流行残差网络 Res101 作为共享卷积层，模型平均检测准确率为 84.44%，平均每张图像的检测时间为 0.184 5 s，相对于传统的 Faster RCNN 模型的不只在准确率有提升，而且还在检测时间上有提升；选用优化后的 Faster RCNN 模型平均检测准确率为 86.50%，平均每张图像的检测时间为 0.138 9 s，不仅时间有提升，检测准确率也有了很大提升。

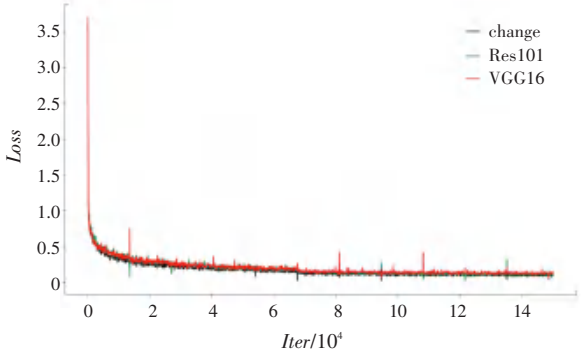

图 4　Loss 结果图

Fig. 4　Loss result diagram

(a) 原图　　(b) 检测后的图

(a) Original figure　　(b) The after detection figure

图 5　测试结果图

Fig. 5　Test result diagram

表 2　测试结果

Tab. 2　Test results

检测方法	平均检测准确率/%	平均检测时间
Faster RCNN(VGG16)	82.85	0.200 2
Faster RCNN(Res101)	84.44	0.184 5
本文模型	86.50	0.138 9

4　结束语

为了对驾驶道路场景图像中的行人和车辆类型

(下转第 100 页)