**Computer Science and Application** 计算机科学与应用, 2022, 12(3), 546-553
Published Online March 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.123055

Hans汉斯

# 基于Faster-RCNN的道路异常状态检测方法研究

**梁　泓，赵曙光**

东华大学信息科学与技术学院，上海

收稿日期：2022年2月8日；录用日期：2022年3月4日；发布日期：2022年3月14日

## 摘　要

**利用无人机等方式开展道路巡检时采集的图像存在图像背景复杂以及目标较小等问题，准确识别道路异常目标成为智能巡检研究热点。本文对两阶段目标检测算法Faster-RCNN进行改进，利用深度残差网络ResNet50作为网络的特征提取backbone，并利用不同层次的特征构造特征金字塔FPN网络，提高了道路异常状态检测模型的性能。**

## 关键词

**道路异常检测，Faster-RCNN，特征金字塔，目标检测**

# Research on Road Abnormal State Detection Method Based on Faster-RCNN

**Hong Liang, Shuguang Zhao**

College of Information Science and Technology, Donghua University, Shanghai

Received: Feb. 8th, 2022; accepted: Mar. 4th, 2022; published: Mar. 14th, 2022

## Abstract

**The images collected when using drones and other means to carry out road inspections have problems such as complex image backgrounds and small targets. Accurately identifying abnormal road targets has become a research hotspot in intelligent inspection. This paper improves the two-stage target detection algorithm Faster-RCNN, uses the deep residual network ResNet50 as the feature extraction backbone of the network, and uses different levels of features to construct a feature pyramid FPN network, which improves the performance of the road abnormal state detection model.**

文章引用: 梁泓, 赵曙光. 基于 Faster-RCNN 的道路异常状态检测方法研究[J]. 计算机科学与应用, 2022, 12(3): 546-553. DOI: 10.12677/csa.2022.123055

**Keywords**

**Road Anomaly Detection, Faster-RCNN, Feature Pyramid, Object Detection**



## 1. 引言

道路在长期的交通负荷以及恶劣的自然环境条件下出现裂缝、坑槽、落石、塌陷等异常状态[1]，传统的道路异常巡检方式是采用人工方式，人工方式需要工作人员到道路现场进行巡视、测量、记录和分析道路问题，该方式受人员主观观念影响较大，检测效率低，危险性较高，实现过程消耗成本大，因而逐渐被淘汰。近年来，深度学习得益于计算机硬件技术的发展，在人工智能、目标检测、图像识别等领域取得了重大进展，卷积神经网络强大的特征提取能力使得计算机视觉任务的精度和效率都达到了极高的水平。基于深度学习的目标检测[2]是指从输入图像中识别出感兴趣目标并将目标类别以及在图像中的位置作为结果进行返回的技术。根据目标检测实现步骤的不同，深度学习目标检测框架可分为基于候选区域的 Two-stage 框架以及基于边框回归的 One-stage 框架。基于候选区域的目标检测算法有：R-CNN [2]、Fast R-CNN [3]、Faster R-CNN [4]等。基于边框回归的检测算法有：YOLO [5]系列、SSD [6]等。

相较于可以直接对输入图像目标检测输出的 One-stage 模型，Two-stage 模型在检测速度上稍逊色，但其检测精准度很高，本文以基于 Two-stage 模型的 Faster R-CNN 算法作为道路异常检测的基础框架，改进其特征提取骨干网络，并引入 FPN 特征融合模型，提高了模型的多尺度特征提取和整合能力，经实验表明，改进后的算法对各类复杂背景下的道路异常情况具有较好的检测和识别能力。

## 2. Faster R-CNN 目标检测算法原理

基于候选区域的 Faster R-CNN 目标检测算法将目标检测分两步完成：首先进行可能包含目标的区域提取并对推荐区域进行特征提取，而后进行目标分类及边框回归得到检测目标的类别及位置。其网络结构如图 1 所示，该模型由特征提取骨干网络(backbone)、区域推荐网络(RPN)、感兴趣区域池化(ROIpooling)层及检测子网络四部分构成，其实现目标检测流程如下：

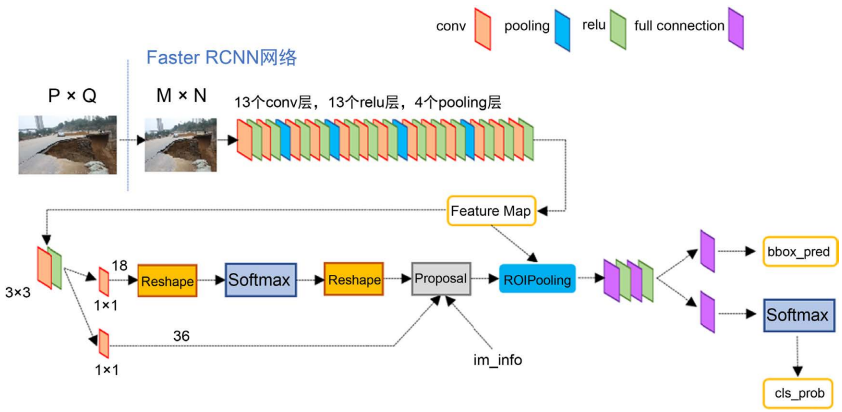


**Figure 1.** Block diagram of Faster R-CNN algorithm
**图 1.** Faster R-CNN 算法框图

1) 首先对输入图像进行尺寸调整，并送入特征提取骨干网络得到图片的特征信息；

2) 将(1)中提取的特征数据输入 RPN 网络，进行是否包含目标判断，并将其位置进行初步框定，生成包含目标信息的推荐区域；

3) 将(2)中生成的推荐区域映射到(1)中的原始特征图上，并进行池化操作得到统一尺寸的推荐框；

4) 对最终的推荐框进行全连接计算，分别进行分类和回归操作，得到目标类别和边框位置。

## 3. Faster R-CNN 算法改进

### 3.1. Backbone 改进

传统的 Faster R-CNN 算法采用 VGG-16 网络[7]作为特征提取 backbone 并将该网络最后一层卷积结果作为提取特征，VGG-16 网络在进行特征提取过程中，池化层的存在回导致特征图尺寸逐渐减小，不利于小目标检测。较深的网络层数有助于提高视觉任务的模型性能，但随着深度神经网络的加深，网络性能会呈现出“退化”现象，并伴随着梯度信号的逐渐减弱[8]。针对上述问题，本文使用具有残差模块的残差网络 ResNet50 (Residual Networks)替换原始的 VGG-16 网络，残差模块是一种通过跳跃连接实现恒等映射的网络结构，如图 2 所示，其中 $x$ 为上阶段网络的输出，$\mathcal{F}(x)$ 为残差映射，通过跳跃连接的 $x$ 为恒等映射，残差块的输出为 $\mathcal{F}(x)+x$ 经激活函数 Relu 处理后的结果。当出现梯度消失情况时，网络由恒等映射得到输出结果。

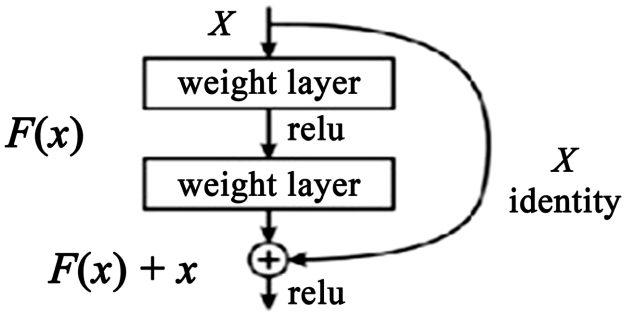


**Figure 2.** Residual structure diagram
**图 2.** 残差结构图

ResNet50 结构如图 3 所示，其由四个阶段的 16 个三卷积层残差块以及首端的卷积层和网络末端的全连接层共 50 层构成，该网络具有网络层次较深且复杂度不高的优点，提高了图片特征提取性能。

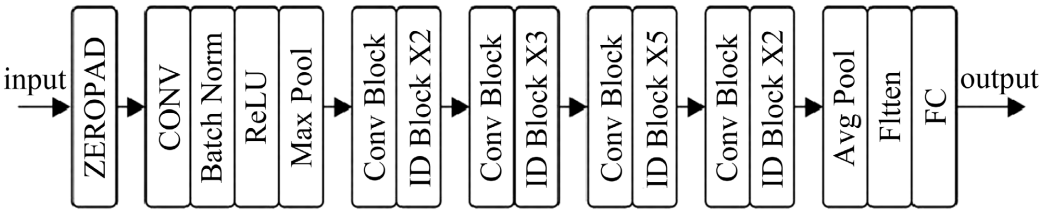


**Figure 3.** ResNet50 network structure
**图 3.** ResNet50 网络结构

### 3.2. 特征金字塔网络

传统的 Faster R-CNN 目标检测算法通常将特征提取网络的最后一层输出作为候选区域推荐网络的输入，网络高层的特征缺乏图片的边缘、纹理等细节信息，不利于进行后续目标定位和小目标检测。在本

文中，由于输入数据集包含航拍图像，目标在图像中占比较小，为了提高模型的检测性能，引入能够对不同尺度层次的特征进行融合的特征金字塔结构(Feature Pyramid Networks, FPN)，增强模型对具有多尺度类型目标的检测能力。为了利用不同尺度的特征信息，FPN 由自顶向下、自下而上以及横向连接几部分组成，其中，自下而上部分，即特征提取网络的正向传播过程，提取各阶段最后卷积层的输出特征图作为该尺度下的特征数据；自上而下部分则对上层特征进行上采样并与邻近的前一阶段卷积层输出进行融合，融合后的各层次的特征都包含丰富的语义和图像细节信息，提高模型对不同尺度图像目标的识别和定位能力。

模型的特征提取及融合如图 4 所示，本文所用特征提取网络为 ResNet50，将该网络的第 2、3、4、5 个卷积块输出的不同尺度特征表示为 C2、C3、C4、C5。特征金字塔在进行特征融合时，首先对最高层特征 C5 进行通道降维，即利用一个 1 × 1 卷积将其通道数变成 256，得到特征层 CP5，随后对其进行上采样，同时对 C4 特征层进行降维操作，最后将 CP5 和降维后的 C4 进行合并操作，并通过 3 × 3 卷积得到融合的特征图 P4。其他尺度特征层的融合操作均按照上述流程完成。最高层特征由于层次最深，直接对其进行降维后经过一个 3 × 3 卷积得到该尺度下的特征图。经 FPN 融合后的多层特征图分别送至 RPN 网络进行候选区域推荐及 ROI 池化层进行建议框截取，为后续目标分类预测和位置回归操作提供多尺度候选框信息。

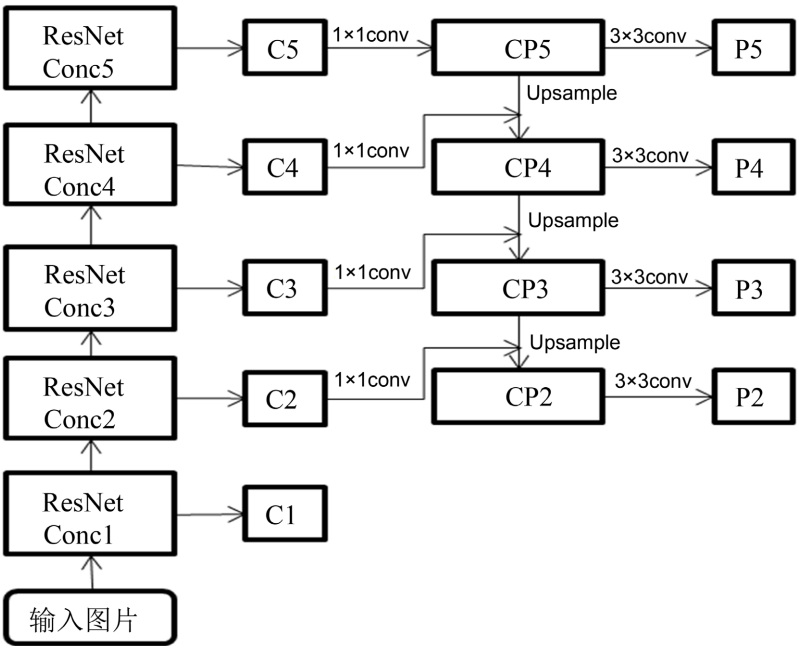


**Figure 4.** Multi-scale feature extraction and fusion structure diagram
**图 4.** 多尺度特征提取及融合结构图

## 4. 数据集准备

本次实验所用数据集通过网络搜索以及实地拍摄取样的方式获得，所涵盖的道路异常情况包括道路塌陷、坑槽、落石以及裂缝四种类型，每种类型共有 500 余张原始图片，为了增加训练数据集，采用旋转、缩放、颜色抖动、水平翻转等方式进行数据增强，最终整理得到道路异常数据集共 8428 张。为了将数据集送入模型进行训练，按照 Pascal VOC 数据集格式对数据集进行整理并使用 Labelimg 图像标注工具对每张图像中的异常类别及边框进行手工标注，并按照 8:2 的比例对数据集进行训练集和测试集划分。

## 5. 实验过程及结论

### 5.1. 实验环境及参数

本次实验所用环境为 Ubuntu 系统，显卡为 Nvidia GeForce GTX2080Ti，内存 11G，深度学习框架为 Pytorch1.6+cuda10.0，采用随机梯度下降算法 SGD 实现模型的反向传播优化，初始学习率设置为 0.0001，动量参数设置为 0.9，权值衰减参数为 0.0005，训练 Epoch 设置为 100。

在深度学习算法中，模型不断求解各个权重参数的最优解，使得模型拟合出的结果与真实情况更加接近，随机梯度下降算法通过对模型的权重参数不断求取偏导数从而确定参数的最优化方向。其算式如(4-1)所示，其中 $\alpha$ 为模型学习率。

$$\omega = \omega - \alpha * \frac{\partial J(\omega)}{\partial(\omega)} \tag{4-1}$$

训练过程分为冻结训练和解冻训练两部分，即在前 50 轮训练中，冻结主干特整提取网络的权重参数，利用在 ImageNet 数据集上的预训练参数作为 backbone 的初始参数对 RPN 网络以及 Fast RCNN 网络进行训练；而在后 50 轮训练中，对整个改进后的网络参数都进行训练，得到更加精确的模型。训练过程中，训练损失及验证损失如图 5 所示，模型在训练过程中，训练和验证损失下降趋势逐渐平缓，即模型逐渐达到收敛状态。

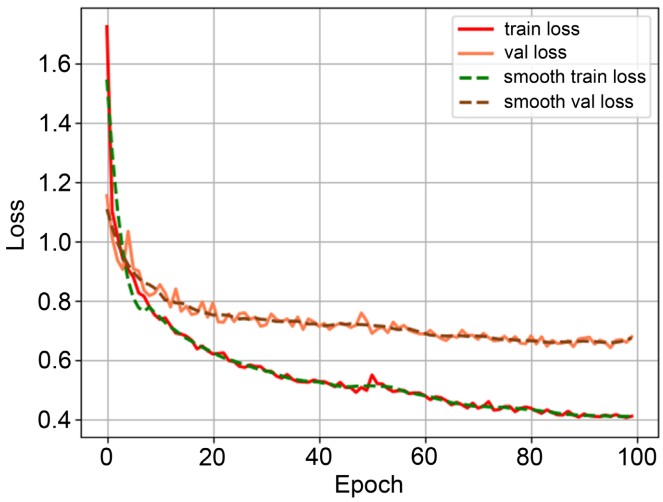


**Figure 5.** Loss changes during training and validation
**图 5.** 训练过程及验证过程损失变化

### 5.2. 实验评估及结论

目标检测算法评价指标通常包含 IoU、FPS、AP 以及 mAP。交并比 IoU 代表算法预测框与真实标注框的重叠程度，其计算公式如(4-2)所示，其中 $B_g$ 代表物体的实际标注边框，$B_p$ 代表经过模型计算所得的预测边框，可以给交并比 IoU 设置一定阈值，用于判断模型预测的边框是否真实包含目标物体。

$$\mathrm{IoU}\left(B_p, B_g\right) = \frac{B_p \cap B_g}{B_p \cup B_g} \tag{4-2}$$

检测速度 FPS，即模型每秒钟所能检测的图片数量，用于对模型速度进行评估。

平均精度 AP (Average Precision)以及平均精度值 mAP (mean Average Precision)是对模型检测精度的

衡量参数，其中 AP 指标代表各类别目标的检测精度，通常由 P-R 曲线进行积分得到，其中 P 为精确率 Precision，其计算公式如式(4-3)所示：

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4-3}$$

式中，TP 代表正样本被真实预测为正样本的个数，FP 代表负样本被错误预测为正样本个数。

R 为召回率 Recall，其计算公式如式(4-4)所示：

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4-4}$$

式中，FN 代表正样本被错误预测为负样本的数量，TP+FN 代表所有正样本数量，召回率通常用于表征模型将正样本能否全部识别出来的能力。

平均精度值 mAP 为模型对各类别检测精度 AP 的平均值，代表模型的整体检测精度。

本文所采用的模型评价标准为平均精度值 mAP，mAP 通过各类检测目标的检测平均精度 AP 进行平均计算后得到。在实验中进行了 3 组对比实验：即模型的 backbone 分别为 VGG-16、ResNet50 以及本文的 ResNet50+FPN，实验结果如表 1 所示，可见，本文算法的检测精度最高，mAP 值达到了 98.82%，采用具有深度残差模块 ResNet50 作为特征提取网络的算法因其具备更深度网络结构，检测精度高于原始的以 VGG-16 为 backbone 的模型。

**Table 1.** Comparison of results of different backbone models
**表 1.** 不同 backbone 模型结果对比

| Model | Backbone | mAP/% |
|---|---|---|
| Faster RCNN | Vgg-16 | 93.4 |
| Faster RCNN | Resnet50 | 97.3 |
| Faster RCNN | Resnet50+FPN | 98.8 |

模型检测效果如图 6~9 所示，从检测效果图中可以看出，改进后的模型对各类道路异常情况实现了准确的标注识别，提高了道路检测的效率和准确度，为实现智能化的道路巡检提供了一种实用便捷的方法。

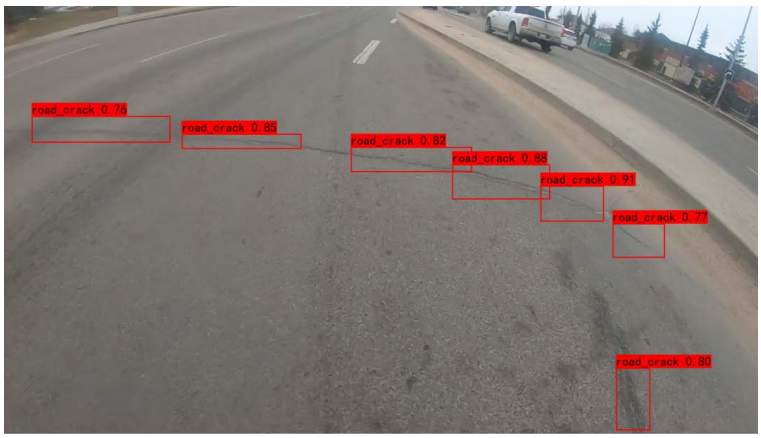


**Figure 6.** The effect of road crack detection
**图 6.** 道路裂缝检测效果图


**Figure 7.** The effect of road pothole detection
**图 7.** 道路坑槽检测效果图


**Figure 8.** The effect of road collapse detection
**图 8.** 道路塌陷检测效果图


**Figure 9.** The effect of road rockfall detection
**图 9.** 道路落石检测效果图

## 6. 结论

本文针对传统的 Faster-RCNN 目标检测网络进行了相应的改进，首先将 VGG-16 特征提取网络替换为网络层次更深、特征提取能力更强的 ResNet50 网络，增强了模型对输入图像的特征提取能力。同时，为了提高模型对复杂道路图像以及小目标异常情况的检测性能，引入特征金字塔模型对不同尺度

的图片特征进行融合，增加了不同层次特征的语义和图片细节信息。经实验证明，改进后的 Faster-RCNN 模型能够精准识别输入图像中的道路异常状态，为智能化道路巡检提供了一种可用于实际场景的检测方法。

## 参考文献


[1] 郑晓光, 杨群, 吕伟民. 沥青路面水损害的病害特征与机理分析[J]. 公路工程, 2006, 31(2): 96-98.

[2] Girshick, R., Donahue, J., Darrell, T., *et al.* (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. arXiv:1311.2524 [cs.CV] https://doi.org/10.1109/CVPR.2014.81

[3] Girshick, R. (2015) Fast R-CNN. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1440-1448. https://doi.org/10.1109/ICCV.2015.169

[4] Ren, S., He, K., Girshick, R. and Sun, J. (2016) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031

[5] Redmon, J., Divvala, S., Girshick, R., *et al.* (2016) You Only Look Once: Unified，Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, NV, 27-30 June 2016, 779-788. https://doi.org/10.1109/CVPR.2016.91

[6] Liu, W., Anguelov, D., Erhan, D., *et al.* (2016) SSD: Single Shot Multibox Detector. arXiv:1512.02325 [cs.CV] https://doi.org/10.1007/978-3-319-46448-0_2

[7] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:1409.1556 [cs.CV]

[8] Bengio, Y., Simard, P. and Frasconi, P. (1994) Learning Long-Term Dependencies with Gradient Descent Is Difficult. *IEEE Transactions on Neural Networks*, **5**, 157-166. https://doi.org/10.1109/72.279181