



DOI: 10.12086/oe.2020.190161

融合目标检测与距离阈值模型的露天矿行车障碍预警

卢才武, 齐凡*, 阮顺领

西安建筑科技大学管理学院, 陕西 西安 710055



摘要: 针对当前行车预警方法无法适应露天矿非结构化道路问题, 本文提出一种融合目标检测和障碍距离阈值的预警方法。首先根据露天矿障碍特点改进原有的 Mask R-CNN 检测框架, 在骨架网络中引入扩张卷积, 在不缩小特征图的情况下扩大感受野范围保证较大目标的检测精度。然后, 根据目标检测结果构建线性距离因子, 表征障碍物在输入图像中的深度信息, 并建立 SVM 预警模型。最后为了保证预警模型的泛化能力采用迁移学习的方法, 在 COCO 数据集中对网络进行预训练, 在文中实地采集的数据集中训练 C_5 阶段和检测层。实验结果表明, 本文方法在实地数据检测中精确率达到 98.47%, 召回率为 97.56%, 人工设计的线性距离因子对 SVM 预警模型有良好的适应性。

关键词: 障碍预警; 目标检测; 距离阈值模型; 扩张卷积; 迁移学习

中图分类号: TP391.41

文献标志码: A

引用格式: 卢才武, 齐凡, 阮顺领. 融合目标检测与距离阈值模型的露天矿行车障碍预警[J]. 光电工程, 2020, 47(1): 190161

An open-pit mine roadway obstacle warning method integrating the object detection and distance threshold model

Lu Caiwu, Qi Fan*, Ruan Shunling

School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China

Abstract: In order to solve the problem that the current driving warning method cannot adapt to the unstructured road in open-pit mine, this paper proposes an early warning method that integrates target detection and obstacle distance threshold. Firstly, the original Mask R-CNN detection framework was improved according to the characteristics of open-pit mine obstacles, and dilated convolution was introduced into the framework network to expand the receptive field range without reducing the feature map to ensure the detection accuracy of larger targets. Then, a linear distance factor was constructed based on the target detection results to represent the depth information of obstacles in the input image, and an SVM warning model was established. Finally, in order to ensure the generalization ability of the warning model, transfer learning method was adopted to carry out pre-training of the network in COCO data set, and both the C_5 stage and detection layer were trained in the data collected in the field. The experimental results show that the accuracy and recall of the proposed method reach 98.47% and 97.56% in the field data detection, respectively, and the manually designed linear distance factor has a good adaptability to the

收稿日期: 2019-04-07; 收到修改稿日期: 2019-07-23

基金项目: 国家安全生产重大事故防治关键技术科技项目(0020-2018AQ); 陕西省教育厅专项计划项目(17JK0425)

作者简介: 卢才武(1965-), 男, 博士, 教授, 博士生导师, 主要从事智慧矿山、机器学习相关研究。E-mail: xjdkclu@163.com

通信作者: 齐凡(1995-), 男, 硕士研究生, 主要从事图像检测与识别、深度学习研究。E-mail: XUATqifan@126.com

版权所有©2020 中国科学院光电技术研究所

SVM warning model.

Keywords: obstacle warning; target detection; distance threshold model; dilated convolution; transfer learning

Citation: Lu C W, Qi F, Ruan S L. An open-pit mine roadway obstacle warning method integrating the object detection and distance threshold model[J]. *Opto-Electronic Engineering*, 2020, 47(1): 190161

1 引言

高级辅助驾驶系统(Advanced driver assistance system, ADAS)是行车安全与自动驾驶领域的研究热点。行车防撞预警系统(forward collision warning system, FCWS)根据外部传感器设备感知障碍预警,是 ADAS 的子系统。目前 FCWS 获取碰撞障碍目标信息多采用视觉传感器,相关研究逐渐从传统视觉算法趋向深度神经网络,但大多集中在结构化道路。其特点是道路边界线明显,获取障碍的横向位置和深度信息相对简单。针对露天矿矿区强非结构化道路行车预警问题,还没有明确的数据集与研究成果,研究露天矿行车预警对拓宽机器视觉预警领域的应用范围,保证矿区行人与驾驶员安全有着重要意义。

基于机器视觉的障碍预警研究中,可以分为障碍物检测与目标距离信息获取两阶段。障碍物检测一直是计算机视觉领域的重要课题,传统的经典方法有 Dalal 等^[1]提出的 HOG 特征结合分类器, Felzenszwalb 等^[2]设计的可变形部件模型(deformable part model, DPM)。随着卷积神经网络的提出,以 R-CNN^[3](region convolutional neural networks)为开端的深度学习检测方法性能大幅度超越传统的视觉算法。Fast R-CNN^[4]与 Faster R-CNN^[5]在 R-CNN 的基础上分别提出锚点定位思想和内嵌的区域建议网络使得目标定位的效率和精确度都进一步提高,形成了 two-stage 的网络框架模型。但 two-stage 网络的时效性一直较低,随后有学者提出将边框回归与分类相结合的 one-stage 网络,经典的方法有 yolo^[6]与 yolo9000^[7],但由于联合损失函数和人为设定划分边框等因素,检测精度受数据集影响波动较大,对小目标检测困难。随后 Liu 等^[8]结合区域建议网络和锚点思想提出 SSD(single shot multibox detector)网络。Joseph Redmon 将原先 yolo9000 中的 softmax 改成 logistic 损失并且引入带有残差连接的骨架网络,提出了 yolov3,精度非常逼近 two-stage 网络,且时效性超越所有的 two-stage 网络,是当前最优异的目标检测网络之一。2017 年,何恺文等^[9]首次将实例分割和目标检测相互结合提出 Mask R-CNN,在改进 Faster R-CNN 后加入 FPN^[10](feature pyramid networks)

网络,提高了对遮挡目标和小目标的检测与分割精度。

在目标距离估计方面,结构化道路预警中多采用单目相机标定几何测距的距离估算方法,在路面状况良好的结构化道路上可以有效控制误差,且计算时间短^[11-12]。但在矿区采场中,道路没有明确的车道线、道路边缘与路宽标准,坡度大弯道较多,单目测距的误差较大。当前基于机器视觉的深度估计研究多集中在像素级别,其中无监督估计^[13]的时效性差,无法保证预警的实时性要求;有监督估计^[14]需要大量的深度标注样本。融合像素级深度估计和目标检测的多任务网络难以实现底层的权值共享,而单任务级联组成的网络无法满足实时预警的需求。Teichmann 等将语义分割和目标检测网络融合,提出 MultiNet^[15],实现了道路可行域判定和障碍检测统一的网络框架,在 KITTI 数据集(1248 pixels×384 pixels)上分别以 VGG16、ResNet50 与 ResNet101 为骨架网络,均有良好的精确率,其检测速度为 42.48 ms、60.22 ms、79.70 ms。Xu 等^[16]在此基础上提出 HCNet(highly coupled network)进一步提高了检测速度,但仅仅将路面分割与行人检测融合无法满足预警需求,障碍物深度信息仍是行车预警最关键的信息之一。Chen 等^[17]融合 Faster R-CNN、SSD,提出了集语义分割、深度估计与目标检测的多任务网络 DSPNet(driving scene perception network),首次引入实例深度估计概念。主要思想是将检测实例的距离信息加入原先的目标检测解码层网络,中心点分支同时输出距离信息,使目标检测和深度估计共享权重,避免了昂贵的像素级深度估计。但网络将实例级深度估计放在了目标检测解码层,采用回归损失与原先的目标检测损失相加,很难平衡两者之间的梯度传播效应,影响了检测精度。从实验结果来看,对部分距离很近的车辆检测困难,难以实现闯入式危险源的预警。彭秋辰等^[18]在 Mask R-CNN 基础上引入双目视差测距,在 KITTI 数据集上的实验,对大于 10 m 的目标误差很大,在 GTX1080 实验平台检测时间为 218 ms,较难满足行车预警的精度和时效性。

在道路消失点检测^[19-20]与深度估计中普遍认为障

碍成像大小与行车距离之间成线性关系。驾驶员在判断障碍物距离时不会直接计算与行车的绝对距离，通常是结合经验等因素对危险距离做出模糊判断形成危险距离的视觉阈值^[1]。根据上述思想，针对深度估计与目标检测难以融合问题，本文改进 Mask R-CNN，将扩展卷积引入原先的骨架网络中，并在目标检测与分割后设计线性距离因子，建立支持向量机预警模型。有监督地对危险距离进行识别，避免视觉信息到实际距离估计然后又形成危险距离阈值判断的过程，对道路状况复杂、封闭、低速的露天矿行车有良好适用性。

本文的贡献点如下：1) 改进 Mask R-CNN 的骨架网络，将扩展卷积加入 C_5 阶段，固定特征图大小增大视觉感受野，防止语义瓶颈增强对预警目标的检测精度。2) 在检测基础上提出行车障碍预警模型。避免计算量巨大的深度估计过程，可以有效地对露天矿矿区道路行车障碍进行识别和预警。3) 由于当前自动驾驶道路没有针对矿区非结构化道路的图像数据库，自行采集和标注了针对障碍检测和危险距离阈值预警的露天矿行车预警数据集。

2 目标检测框架

2.1 Mask R-CNN

Mask R-CNN 将 FPN 与 Faster R-CNN 相结合，成为集实例分割与目标检测于一体的深度神经网络模型，其集成效果比单任务网络更优秀。Mask R-CNN 结构如图 1 虚线框所示，主要由骨架网络、上采样层、区域建议网络、ROIAlign、检测分割层组成。

骨架网络是由多层卷积层和池化层组成，经过卷积提取可以将图像的底层信息转化为语义信息更丰富尺寸更小的特征图。通常骨架网络的性能决定后续网

络任务的执行效果，对检测和分割精度有很大影响，原先网络采用深度残差网络(ResNet50)^[21]。

上采样阶段，取 ResNet50 后四个阶段的特征图为输入特征。将 C_5, C_4, C_3 上采样分别与 1×1 卷积后的 C_4, C_3, C_2 相加得到 P_4, P_3, P_2 ，而 P_5 由 C_5 直接卷积所得。为了消除叠加效应，将 P_2, P_3, P_4, P_5 再做卷积形成大小固定的特征图，同时将 P_5 做最大池化得到 P_6 。由于 P_6 的语义信息非常集中，因此只在区域建议网络中生成 Anchor 信息，而不直接进行 Pooling 处理。上述过程用式表示如下：

$$\begin{cases} P'_i = \text{sum}(\text{upsample}(P'_{i+1}), \text{conv}(C)) \\ P_6 = \text{maxpooling}(P_5) \\ P_i = \text{conv}(P'_i) \end{cases}, \quad (1)$$

式中： $i=5,4,3,2$ ，当 i 为 5 时 $\text{upsample}(P'_{i+1})$ 为 0。sum、upsample、maxpooling 分别代表对位直接相加求和，最邻近插值上采样和步长为 2 的最大值池化。

区域建议网络将得到的所有特征层作为输入。首先根据预设的长宽比和区域面积大小遍历所有的特征图的像素点生成 Anchor。由于矿区行车属于封闭环境，行车障碍主要为前车和行人，站姿行人与前车长宽比集中在 3~5 与 1~2 之间。因此将长宽比设定为 [3, 1, 0.5]，符合障碍目标的图像特征。然后将输入的特征图进行卷积操作得到 $\Delta x, \Delta y, \Delta h, \Delta w$ 和前后置信度 (l)，最后将排除与真实 box 交并比小于 0.7 的 Anchor，并对中心坐标和长宽进行修正。计算式：

$$\begin{cases} x = (1 + \Delta x) \cdot x \\ y = (1 + \Delta y) \cdot y \\ w = \exp(\Delta w) \cdot w \\ h = \exp(\Delta h) \cdot h \end{cases}. \quad (2)$$

在进入 ROIAlign 之前，将所有保留的 Anchor 进行 NMS(non-maximum suppression)算法过滤，保留与真实 box 接近的候选框去掉冗余候选框，文中设定 NMS 的最大保留框为 300 个。ROIAlign 是将生成的 Anchor 采用双线性插值的方法对应到特征图上进行裁剪，为后续分割和检测提供输入。在 ROIAlign 之后将生成的特征图分别输入目标检测和分割网络，检测网络由全连接层和 softmax 层组成，用于对前景物体的精确分类。分割网络经过多个同尺度卷积和两个反卷积层将输出固定，它代表候选框某个类别输出掩模对应原图的前后置信度。

2.2 Block 单元改进

由于预警目标通常偏向图像中央且像素占比较

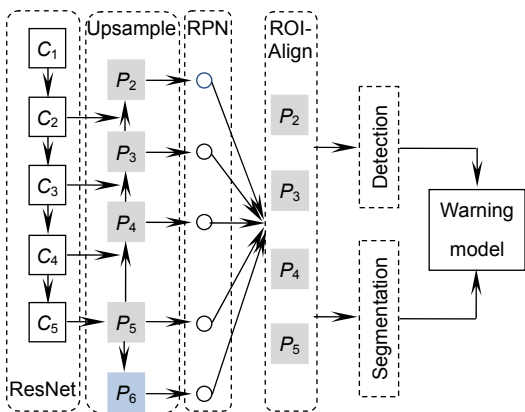


图 1 改进 Mask R-CNN 框架

Fig. 1 Improved mask R-CNN framework

大，所以网络需要尽可能提高大目标的定位与分割精度。Mask R-CNN 的骨架网络 ResNet 分为 5 个阶段，除第一阶段外，其余阶段都会缩小特征图的大小。Li 等^[22]认为过大的下采样导致大物体边界区域的检测和定位精度下降，但不缩小特征图尺寸影响高层卷次特征的感受野，出现语义瓶颈不易于检测和分割。因此本文引入扩展卷积^[23](dilated convolution)Block 单元，将 stage5 中 Block 的 3×3 卷积替换为 3×3 的空洞卷积，并且固定特征图与 stage4 相同，但增加感受野范围。扩展卷积如图 2 所示。

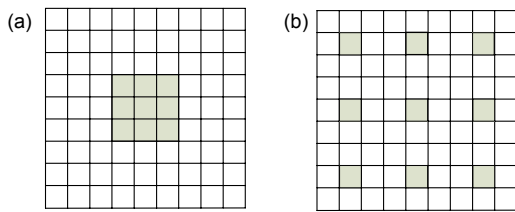


图 2 两种卷积操作。

(a) 常规卷积; (b) 扩展率为 2 的扩展卷积

Fig. 2 Two convolution operations.

(a) A conventional convolution;
(b) An dilated convolution with

从图 2 可以发现，常规 3×3 卷积操作的感受野为 3×3，扩展率为 2 的扩展卷积为 7×7，在参数量相同的情况下感受野扩大一倍。扩展卷积的扩展率越大感受野也就越大，对较大目标的分类和分割精度会提高。

图 3 为改进 C₅ 阶段的连接图，其中保持原有 ResNet 中 Block 子网结构不变，将中间层卷积替换为扩展卷积，子网底部采用对位像素相加的残差连接。改进 C₅ 阶段在残差连接的基础上结合了扩展卷积的非局部感受野，解决了梯度消失以及局部连接只能通过收缩特征图增大感受野的问题。对于露天矿预警阈值目标来说，一般为中等目标，像素区域小于 64 的目标距离非常远，对行车安全的影响不大，因此上述结构综合考虑实际任务特性。

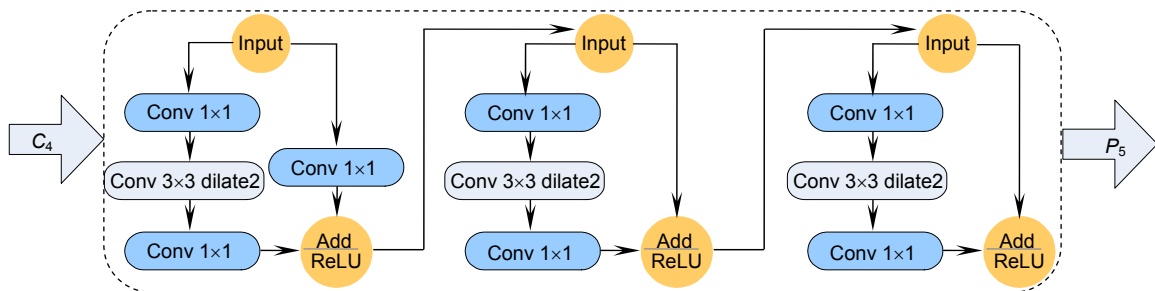


图 3 C₅ 阶段引入空洞卷积

Fig. 3 Empty convolution is introduced in C₅

3 预警模型

人眼在驾驶预警时不会精确估计距离，通常是形成模糊判断形成危险距离的阈值。在数字图像中，障碍成像大小与行车距离之间成线性关系，障碍物类别以及姿态确定的情况下行车与前方障碍的纵向关系通过障碍物图像的大小来表达^[11]。据此，文中对预警特征进行人工设计，将输出图像检测框的长、宽、面积以及掩模面积进行归一化处理，结合检测类别(*l*)建立线性距离因子作为预警模型的特征。具体设计如下：

$$\begin{cases} w_{\text{anchor}} = w'_{\text{anchor}} / w \\ h_{\text{anchor}} = h'_{\text{anchor}} / h \\ s_{\text{anchor}} = s'_{\text{anchor}} / s \\ s_{\text{mask}} = s'_{\text{mask}} / s \end{cases}, \quad (3)$$

式中： w_{anchor} 、 h_{anchor} 、 s_{anchor} 分别代表归一化后检测框的宽、高和面积， s_{mask} 代表分割掩模的面积与总面积的比值， w 、 h 和 s 则代表输入图像的相应尺寸。在类别确定时，上述因子可以将障碍目标的深度信息转化为视觉范围线性空间中的比例属性。相比 Chen^[17]将深度信息标注后共享卷积层的做法，文中没有预测障碍物距离而是直接建立预警模型分别训练检测网络与预警模型，避免多任务网络损失难以平衡影响检测精度的问题。为保证后续预警分类的时效性，实验中选取经过目标检测后置信度最高的边框与掩模。

如上所述，实验中将行车障碍分为预警目标与安全目标，两者在宽、高、面积的数据分布有明显差异，不满足线性可分条件。针对预警与非预警的多维二分类问题，SVM 具有很好的适用性。由于分类特征的维度较低进行核映射的计算消耗较少，因此文中选择高斯核 SVM 模型进行预警判断，模型具体表述如下：

$$\begin{cases} \min \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^N \alpha_i \\ \text{s.t.} \sum_{i=1}^N \alpha_i y_i, \quad 0 \leq \alpha_i \leq C, \quad i = 1, 2, \dots, N \end{cases}, \quad (4)$$

其中：特征集为

$$T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\},$$

$$x_i = (w_{\text{anchor}}, h_{\text{anchor}}, s_{\text{anchor}}, s_{\text{mask}}, l), y_i = (0 \ 1),$$

核函数 $K(x_i, x_j) = \exp(-(x_i - x_j)^2 / 2g)$ ， α_i 为拉格朗日乘子， C 为惩罚值， g 为核参数。惩罚值和核参数相互独立均影响模型的精度和泛化能力，经实验，文中 g 选择 0.1， C 为 1。

4 实验

4.1 数据采集与模型训练

当前数据集如 KITTI、Cityscapes、BDDV 等均是针对城市道路采集，道路环境与露天矿采场不同，障碍距离较远且偏向结构化，没有采场中的大型运输车量。文中以某大型露天矿采场为主(开采深度小于 100 m，台阶面坡脚约为 70)，采用 Canon EOS 80d 数码相机与激光测距仪采集数据，行车障碍主要为行人与前车。每个采场根据矿坑的开采阶段、路面状况、运载卡车载重等多因素制定安全行驶距离，实地考察了解采场的行车安全距离规定为 30 m，因此当测距仪检测到直线距离小于 30 m 即判定为危险。总共采集 916 张图像，包括直行道路、转弯、对向来车和跟车，以及在不同光照条件下的图像数据。由于矿区中的障碍目标较为稀疏，数据集中共有 1049 个障碍目标。其中车辆 786 个，行人 263 人，预警目标和安全目标组成如图 4 所示，基本满足样本的均衡性要求。随机选择 687 张作为训练集，其余 229 张为测试集，图像原分辨率是 4032×3024，在输入网络时将图像大小调整为 512×512。

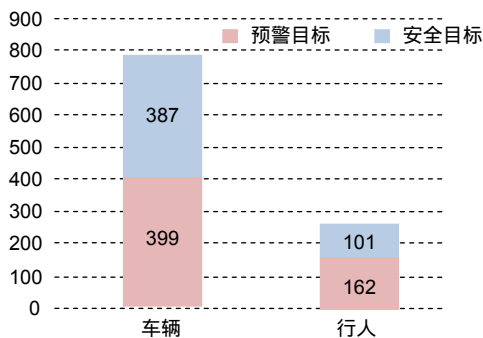


图 4 数据集组成

Fig. 4 Data set composition

由上述可知，网络采用改进的 ResNet50 作为骨架网络参数较多，仅用 687 张图像训练导致网络过拟合降低泛化能力。为解决这个问题，首先采用迁移学习方法将原网络在 COCO 数据集进行预训练，在此基础

使用采集到的数据集继续训练 C_5 和检测层。同时采用数据增强法，将所有训练图像左右反转并添加半径为 0~5 之间的均匀随机高斯模糊处理，以扩大数据集。

实验配置为 NVIDIA GeForce GTX 1080，在改进 Mask R-CNN 的训练阶段采用小批量 SGD，初始学习率为 0.001，动量参数为 0.9，衰减率为 0.005，迭代 2000 次。当 Mask R-CNN 训练结束后会将所有最高置信度检测框的数据作为 SVM 模型的训练集。

4.2 实验结果与分析

为对比不同特征检测有效性，将提出的五种特征值各自组合分别建立 SVM 分类器进行预警分类，以精确度(Accuracy)、召回率(Recall)、 F_1 三个指标进行模型评价。

$$\text{精确度: } P = \frac{TP}{FP + TP}, \quad (5)$$

$$\text{召回率: } R = \frac{TP}{FP + FN}, \quad (6)$$

$$F_1 \text{ 分数: } F_1 = \frac{2P \cdot R}{P + R}, \quad (7)$$

其中： TP 为分类正确的预警目标数， FP 为分类错误的预警目标数， FN 为分类错误的安全目标数。

可以从表 1 看到，不同特征对分类的精度影响存在差异，其中类别属性的信息增益最大，当加入障碍类别后所有对比项的精确率均达到了 88%以上。这是因为不同类别间的预警基准不同，同种障碍之间的线性因子才能表征距离属性。实验 6 使用五种特征可以获得最好的预警效果，而相比目标检测与分割阶段其预警分类的检测时间短。因此为保证预警精度，采用检测框宽度(w_{anchor})、高度(h_{anchor})、面积(s_{anchor})以及掩模面积(s_{mask})和类别(l)共五种特征作为预警分类依据。

文中将当前最优的 one-stage 网络 yolov3 以及 Mask R-CNN 对比，其中 yolov3 没有实例分割层，其提取的特征为 w_{anchor} 、 h_{anchor} 、 s_{anchor} 、 l ，结果如表 2。可以看到，三种网络框架整体的预警表现均较优异，改进后 Mask R-CNN 的预警检测效果最好，精度为 98.47%，召回率 97.56%。由于参数数量和运算量没有增加，时效性与改进前的网络基本一致。yolov3 网络在小目标检测方面表现突出，时效性好，对所有预警目标召回率较高，但预警分类的精确度由于输入特征缺少掩模面积略有不足。Mask R-CNN 由于 FPN 阶段的下采样步幅太大导致大物体的边界检测精确度不高，但掩模面积增加了输入 SVM 的特征维度从而提高预警模型超平面建立的精确性。文中为了保证 30 m

表 1 不同特征组合对比

Table 1 Comparisons of different feature combinations

Type	Accuracy/%	Recall/%	F_1 /%
1 $W_{anchor}+h_{anchor}+s_{anchor}+l$	88.12	95.13	91.49
2 $W_{anchor}+h_{anchor}+s_{mask}+l$	90.16	92.84	91.48
3 $h_{anchor}+s_{anchor}+s_{mask}+l$	90.22	81.36	85.56
4 $W_{anchor}+s_{anchor}+s_{mask}+l$	94.34	82.64	88.10
5 $W_{anchor}+h_{anchor}+s_{anchor}+s_{mask}$	55.61	64.92	59.51
6 $W_{anchor}+h_{anchor}+s_{anchor}+s_{mask}+l$	98.47	97.56	98.01

表 2 三种检测框架对比

Table 2 Comparisons of three detection frameworks

Model	Accuracy/%	Recall/%	F_1 /%	Time/ms
yolov3+SVM	95.08	95.31	95.19	87
Mask R-CNN+SVM	96.64	95.89	96.26	134
Ours model	98.47	97.56	98.01	136

的阈值范围内的检测精度，对骨干网络取消 32 到 64 的感受野跨度，采用扩展卷积保证高层感受野，有益于对预警模糊目标的定位与特征提取及 SVM 分类。

图 5 为识别预警结果，其中包括跟车、对向来车、直线会车、弯道会车以及遭遇行人等多种露天矿运载卡车行车路况。图 5(a)~5(c)均为安全目标的识别结果，图 5(d)~5(j)中主要为预警目标；红色边框代表识别结果为预警目标，绿色边框则为安全目标。总体来看，三种框架与预警模型结合后均可有效识别安全目标与危险目标。图 5(e)中 yolov3 与 Mask R-CNN 均将目标

分类错误，Mask R-CNN 是因为分割和定位出现了较大偏差，而 yolov3 则是检测框过窄，矿车被错分为安全目标。图 5(i)中视角为侧向近距离会车，目标较大且出现了严重的重叠，Mask-R-CNN 识别出 3 辆卡车，yolov3 无法正确分割粘连目标。文中网络则相对精确地检测出卡车，说明了改进框架对较大目标检测预警的优越性，契合复杂环境下闯入式危险的识别特点。图 5(f)与 5(j)中出现远距离行人和矿车，yolov3 与 Mask-R-CNN 可以较好地检测出这类小目标，而本文网络因为扩展卷积的扩展率问题存在个别微小目标的

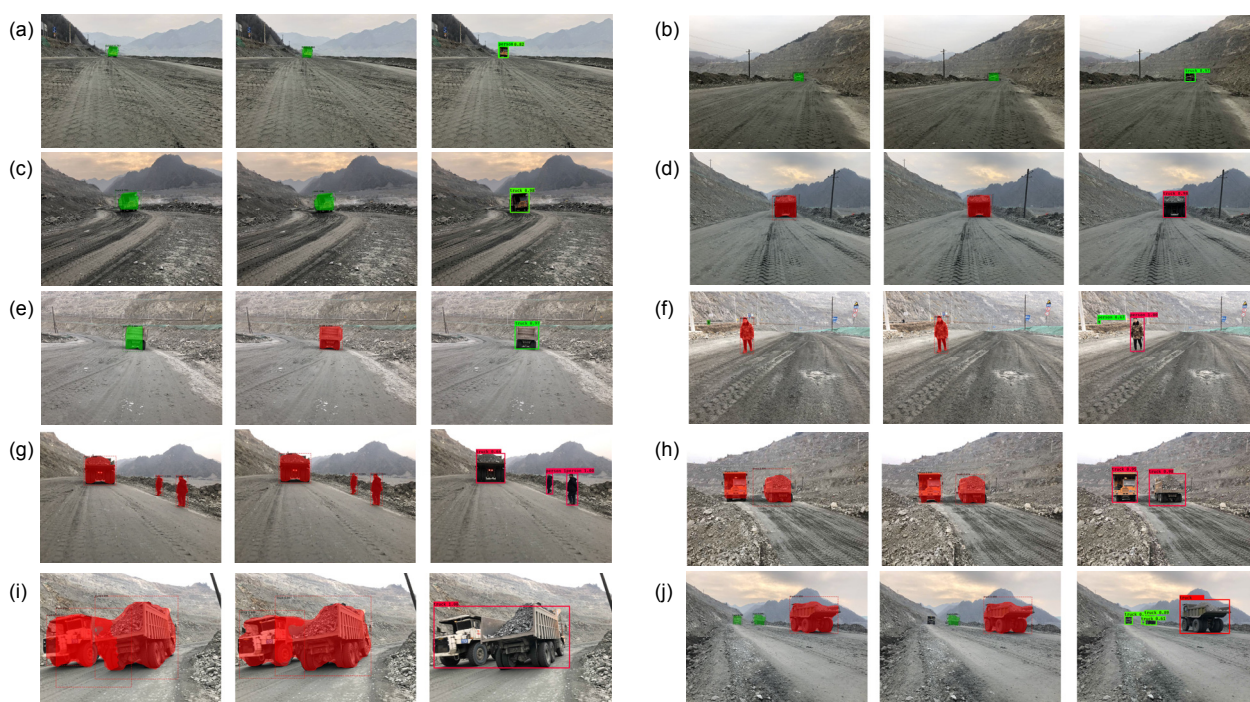


图 5 三种算法在多种场景下的预警效果对比图。从左到右分别是 Mask R-CNN、本文框架、yolov3 经过预警模型分类的检测结果，红色代表检测为预警目标，绿色代表安全目标。

(a) 会车场景一；(b) 会车场景二；(c) 会车场景三；(d) 跟车场景一；(e) 跟车场景二；

(f) 行人场景；(g) 跟车与行人复杂场景；(h) 中距离多车交会；(i) 近距离多车交会；(j) 远距离多车交会

Fig. 5 Comparison diagram of three models in various scenarios. From left to right are the detection results of Mask R-CNN, framework of this paper and yolov3 classified by the warning model. The red represents the detection of warning targets and the green represents the security targets. (a) Meeting scene 1; (b) Meeting scene 2; (c) Meeting scene 3; (d) Following the car scene 1; (e) Scene 2 with the car; (f) Pedestrian scene; (g) Complex scenes with cars and pedestrians; (h) Medium-distance multi-vehicle meeting; (i) Close multiple vehicle meeting; (j) Long distance multi-vehicle meeting

漏检,但通常微小目标对安全行车的影响较小,文中模型设计主要注重大目标的表达精度。

5 结论

本文针对露天矿非结构化道路行车障碍预警,提出了改进 Mask R-CNN 网络框架并人工设计线性距离因子和预警模型。同时针对当前露天矿道路障碍预警缺少有效数据集问题,对实际场景进行设计、取样,组建了适合非结构化障碍检测的矿区行车数据集。对比实验结果表明,文中设计的线性因子可以表达障碍物的距离特征对预警模型有良好的适应性。改进模型的精确度、召回率分别为 98.47%、97.56%,超过原先的 Mask R-CNN 与 yolov3 框架,对距离阈值附近的障碍检测精度较高。文中将深度学习检测框架与预警模型相结合,避免复杂的深度估计实现障碍物预警,证明了目标检测网络在矿区复杂道路检测与预警的可行性,为今后的露天矿行车预警研究提供了思路。但由于数据集实地获取比较困难且危险性较高,所以数据集较小。下一步应该考虑扩展数据集,并进一步提高模型的时效性。

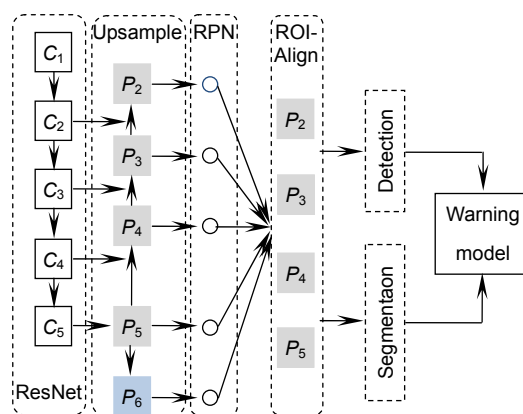
参考文献

- [1] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]// 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2005: 886–893.
- [2] Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model[C]// Proceedings of 2008 IEEE Conference on Computer Vision and Pattern Recognition, 2008: 1–8.
- [3] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580–587.
- [4] Girshick R. Fast R-CNN[C]// Proceedings of 2015 IEEE International Conference on Computer Vision, 2015: 1440–1448.
- [5] 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.
- [6] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779–788.
- [7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517–6525.
- [8] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[C]// The 14th European Conference on Computer Vision, 2016: 21–37.
- [9] He K M, Gkioxari G, Dollár P, et al. Mask R-CNN[C]// Proceedings of 2017 IEEE International Conference on Computer Vision, 2017: 2980–2988.
- [10] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 936–944.
- [11] Yang H C, Zhu W B, Tong Y. Pedestrian collision warning system based on looking-in and looking-out visual information analysis[J]. *CAAI Transactions on Intelligent Systems*, 2019, **14**(4): 756–760.
杨会成, 朱文博, 童英. 基于车内外视觉信息的行人碰撞预警方法[J]. *智能系统学报*, 2019, **14**(4): 756–760.
- [12] Yang D F, Sun F C, Wang S C, et al. Simultaneous estimation of ego-motion and vehicle distance by using a monocular camera[J]. *Science China Information Sciences*, 2014, **57**(5): 1–10.
- [13] Xu Y F, Wang Y, Guo L. Unsupervised ego-motion and dense depth estimation with monocular video[C]// Proceedings of 2018 IEEE 18th International Conference on Communication Technology, 2018: 1306–1310.
- [14] Tateno K, Tombari F, Laina I, et al. CNN-SLAM: real-time dense monocular SLAM with learned depth prediction[C]// Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6565–6574.
- [15] Teichmann M, Weber M, Zöllner M, et al. MultiNet: real-time joint semantic reasoning for autonomous driving[C]// Proceedings of 2018 IEEE Intelligent Vehicles Symposium (IV), 2018: 1013–1020.
- [16] Li B J, Liu S, Xu W C, et al. Real-time object detection and semantic segmentation for autonomous driving[J]. *Proceedings of SPIE*, 2017, **10608**: 106080P.
- [17] Chen L F, Yang Z, Ma J J, et al. Driving scene perception network: real-time joint detection, depth estimation and semantic segmentation[C]// Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision, 2018: 1283–1291.
- [18] Peng Q C, Song Y C. Object recognition and localization based on Mask R-CNN[J]. *Journal of Tsinghua University (Science and Technology)*, 2019, **59**(2): 135–141.
彭秋辰, 宋亦旭. 基于 Mask R-CNN 的物体识别和定位[J]. *清华大学学报(自然科学版)*, 2019, **59**(2): 135–141.
- [19] Kong H, Audibert J Y, Ponce J. Vanishing point detection for road detection[C]// Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009: 96–103.
- [20] Moghadam P, Starzyk J A, Wijesoma W S. Fast vanishing-point detection in unstructured environments[J]. *IEEE Transactions on Image Processing*, 2012, **21**(1): 425–430.
- [21] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770–778.
- [22] Li Z M, Peng C, Yu G, et al. DetNet: design backbone for object detection[C]// The 15th European Conference on Computer Vision, 2018: 339–354.
- [23] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[EB/OL]. (2016-04-30). <https://arxiv.org/abs/1511.07122v2>.

An open-pit mine roadway obstacle warning method integrating the object detection and distance threshold model

Lu Caiwu, Qi Fan*, Ruan Shunling

School of Management, Xi'an University of Architecture and Technology, Xi'an, Shaanxi 710055, China



Improved mask R-CNN framework

Overview: Most of the researches on traffic obstacle warning based on machine vision mainly focus on urban roads. There are no clear road surface, boundary, and road width standard but many steep curves on open-pit mine roads which are quite different from the urban roads. Thus, the target detection and early warning method suitable for urban roads cannot be applied to non-structural open pit roads. With the emergence of convolutional neural network, target detection and depth estimation based on deep learning gradually surpass the traditional computational vision methods in accuracy and applicability. However, target detection and pixel depth estimation are difficult to implement the underlying convolutional layer sharing mechanism. Usually, the driver will make the early-warning judgment based on the distance threshold according to the experience and other factors. Therefore, an early-warning method combining target detection and the distance threshold model is proposed in this paper. First, the original Mask R-CNN detection framework was improved according to the characteristics of open-pit mine obstacles, and the dilated convolution was introduced in ResNet. Under the condition that the subnet structure and residual connection remain unchanged, the 3 by 3 convolution in the C₅ subnet was replaced by the dilated convolution with the dilatation rate of 2, so that the original 3 by 3 receptive field was extended to 7 by 7. The range of the receptive field was expanded to ensure the detection accuracy of the larger target without reducing the feature graph. Then, according to the detection and classification results, the normalized detection frame length, width, area, mask area, and category were used as the depth information to represent the obstacle in the two-dimensional image. Radial basis function SVM warning model was established to judge whether the detected target is a dangerous target. Finally, in order to ensure the generalization ability of the warning model, transfer learning method was adopted to pre-train the network in COCO data, so that sufficient underlying characteristic information was learned in the first four stages. Both C₅ stage and detection layer were trained in the data collected in this paper. The experimental results show that the linear distance factor proposed in this paper can effectively represent the depth information of obstacles, and the Mask R-CNN and yolov3 can adapt to the warning model in this paper. The improved Mask R-CNN in this paper pays more attention to the regression and classification of frames with larger targets, with an accuracy rate of 98.47% and a recall rate of 97.56% which are better than other frames.

Citation: Lu C W, Qi F, Ruan S L. An open-pit mine roadway obstacle warning method integrating the object detection and distance threshold model[J]. *Opto-Electronic Engineering*, 2020, 47(1): 190161

Supported by Technological Projects for Prevention and Control of Severe and Extraordinary Accidents in National Safety Production (0020-2018AQ) and Special Project of Shaanxi Education Department (17JK0425)

* E-mail: XUATqifan@126.com