

本文引用格式：陈冬青,邓泳涛.面向三维钢结构环境的攀爬机器人可行域检测模型[J].自动化与信息工程,2024,45(6):36-40;46.  
CHEN Dongqing, DENG Yongtao. Feasible region detection model for climbing robots in 3D steel structure environment[J]. Automation & Information Engineering, 2024,45(6):36-40;46.

## 面向三维钢结构环境的攀爬机器人可行域检测模型\*

陈冬青<sup>1</sup> 邓泳涛<sup>2</sup>

(1.广州检验检测认证集团有限公司, 广东 广州 511447

2.广东工业大学机电工程学院, 广东 广州 510006)

**摘要:** 可行域检测对于自主移动攀爬机器人理解周围环境, 保障其高效、安全地移动具有重要意义。针对传统语义分割模型参数量大、计算速度慢等问题, 提出一种改进的 DeepLabv3+模型。该模型的主干网络采用轻量级的 MobileNetv2 替代 Xception, 有效减少了模型的参数量并提升了计算效率; 在空洞空间金字塔池化单元中增加了卷积支路并优化了空洞率, 增强了模型对不同尺寸物体的分割能力; 在解码阶段引入全局注意力机制, 增强模型对上下文信息的聚合能力, 进一步提高模型的分割精度。实验结果表明, 改进的 DeepLabv3+模型相较于 U-Net 模型、DeepLabv3+模型, 在攀爬机器人的可行域检测任务上取得了较好的分割效果。

**关键词:** 攀爬机器人; 可行域检测; 三维钢结构; MobileNetv2; 全局注意力机制

中图分类号: TP242.2

文献标志码: A

文章编号: 1674-2605(2024)06-0005-06

DOI: 10.3969/j.issn.1674-2605.2024.06.005

开放获取

## Feasible Region Detection Model for Climbing Robots in 3D Steel Structure Environment

CHEN Dongqing<sup>1</sup> DENG Yongtao<sup>2</sup>

(1.Guangzhou Inspection, Testing and Certification Group Co., Ltd., Guangzhou 511447, China

2.School of Mechanical and Electrical Engineering, Guangdong University of Technology, Guangzhou 510006, China)

**Abstract:** Feasible region detection is of great significance for autonomous mobile climbing robots to understand the surrounding environment and ensure their efficient and safe movement. An improved DeepLabv3+model is proposed to address the issues of large parameter count and slow computation speed in traditional semantic segmentation models. The backbone network of this model uses lightweight MobileNetv2 instead of Xception, effectively reducing the number of parameters and improving the computational efficiency of the model; We added convolutional branches and optimized the hole rate in the pyramid pooling unit of the hollow space, enhancing the model's segmentation ability for objects of different sizes; Introducing a global attention mechanism in the decoding stage enhances the model's ability to aggregate contextual information and further improves the segmentation accuracy of the model. The experimental results show that the improved DeepLabv3+model achieves better segmentation performance in the feasible domain detection task of climbing robots compared to the U-Net model and DeepLabv3+model.

**Keywords:** climbing robots; feasible region detection; 3D steel structure; MobileNetv2; global attention mechanism

### 0 引言

随着海上航运业的快速发展, 港口大型起重机搬运集装箱的需求日益增加, 工作强度越来越大, 对起

重机的检测也越来越重要<sup>[1]</sup>。目前, 起重机的定期检测主要依赖人工模式, 存在安全风险高、效率低等问题。而攀爬机器人可以携带传感器和检测设备, 到达人

力难以触及的高空区域,对起重机的可行域进行检测。

近年来,神经网络在计算机视觉领域取得较大的进展,尤其在图像分类<sup>[2]</sup>、语义分割<sup>[3]</sup>等图像处理方面表现较好。攀爬机器人在起重机这类三维钢结构上的可行域检测是一项典型的语义分割任务,其核心在于先辨识图像中每个像素点的类别特征,再区分任务所需的区域与其他区域<sup>[4]</sup>。相较于卷积神经网络的其他应用,语义分割既要提取图像的具体特征,又要捕捉图像整体的高层语义信息,使任务更加复杂。直到 SHELHAMER<sup>[5]</sup>等提出了全卷积网络 (fully convolutional networks, FCN),才解决了图像尺寸限制的问题,并利用逐步连接融合了不同层级的语义信息;尽管反卷积上采样有助于恢复图像分辨率,但图像细节分割仍有不足。文献[6]提出 U-Net 结构,利用编码器捕获深层特征和高级语义信息,并在解码器中通过逐步连接的方式整合不同尺度的特征,获得更精确的分割结果;但因同时引入了大量参数,降低了模型的效率。文献[7]提出 SegNet 结构,其与 U-Net 结构相似,且在最大池化过程中保存了像素的位置信息,在上采样时能精确还原图像细节;该网络在车道分割任务中表现出色,但在跨领域场景中泛化能力不足。文献[8]提出 RefinetNet,通过长距离残差连接和下采样特征图的融合,提升了分割精度;但对计算和存储的要求较高。Google 提出了一系列的 DeepLab 模型<sup>[9-12]</sup>,但这些模型在处理速度和模型容量方面不具优势。

针对上述问题,本文基于 DeepLabv3+模型,构建一种轻量化的语义分割模型(改进的 DeepLabv3+模型)。该模型的主干网络采用轻量级的 MobileNetv2 替代 Xception,以减少模型的参数量及提升计算效率;在解码阶段引入全局注意力机制(global attention mechanism, GAM),增强模型对上下文信息的聚合能力,以提高模型的分割精度。攀爬机器人通过实时的语义分割来更新可行域,可有效避免机器人误入危险区域而造成跌落事故。

## 1 相关研究

### 1.1 MobileNetv2

MobileNetv2 是一个高效、轻量级的深度学习模型<sup>[13-15]</sup>,它引入了一种倒置残差结构,先进行深度线性处理,再进行逐点卷积,有效地减轻了模型的计算负担。该模型采用深度可分离卷积,进一步降低了模型的复杂性,并通过引入宽度乘数,允许模型根据需求调整大小和计算量,同时能够保证模型性能。此外,MobileNetv2 通过 ReLU 激活函数来增强模型的表达能力,并利用全局平均池化层替换传统的全连接层,进一步减少了模型的参数量。MobileNetv2 的总体架构如表 1 所示。其中,  $t$  为输入通道的乘法系数(即扩展倍数),  $c$  为输出通道的数量,  $n$  为模块的重复次数,  $s$  为卷积步幅,  $k$  为宽度缩放因子。

表 1 MobileNetv2 的总体架构

输入	操作	$t$	$c$	$n$	$s$
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$14^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d $1 \times 1$	-	1 280	1	1
$7^2 \times 1 280$	avgpool $7 \times 7$	-	-	1	-
$1 \times 1 \times 1 280$	conv2d $1 \times 1$	-	$k$	-	-

### 1.2 GAM

GAM 是一种提升深度学习模型特征表达能力的机制<sup>[16]</sup>,适用于需要捕获全局依赖性的任务。GAM 通过对图像或序列数据进行全面分析,强化模型对整体上下文信息的关注;通过动态权重分配和上下文信息聚合,突出全局上下文信息的重要性。GAM 融合了通道注意力机制和空间注意力机制,结构如图 1 所示。

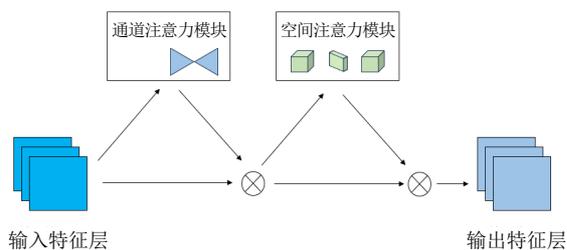


图1 GAM 结构

GAM 处理特征图的流程可通过公式(1)来描述:

$$\begin{cases} F_1 = M_C(F) \otimes F \\ F_2 = M_S(F_1) \otimes F_1 \end{cases} \quad (1)$$

式中:  $F_1$  为经通道注意力模块处理后的特征图,  $M_C$  为通道权重系数,  $\otimes$  为逐元素乘法,  $F_2$  为经空间注意力模块处理后的特征图,  $M_S$  为空间权重系数。

通道注意力模块结构如图 2 所示。

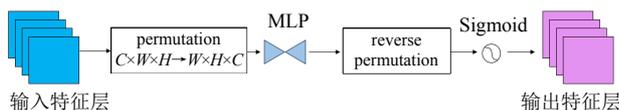


图2 通道注意力模块结构

首先, 对输入特征图进行维度变换; 然后, 将维度变换后的输入送入多层感知机 (multilayer perceptron, MLP), 以增强不同维度之间的通道和空间依赖性 (MLP 包括 2 个线性层和 1 个 ReLU 激活函数: 第一个线性层将特征图按压缩比例  $r$  进行通道压缩, 并利用 ReLU 激活函数减缓梯度消失和爆炸现象; 第二个线性层将特征通道维度还原); 最后, 利用 Sigmoid 函数对输入特征图进行非线性归一化处理。

空间注意力模块结构如图 3 所示。

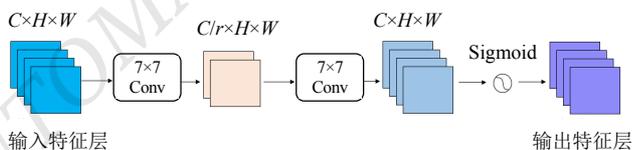


图3 空间注意力模块结构

首先, 接收通道注意力模块输出的特征图; 然后, 通过 2 个  $7 \times 7$  的卷积层进行特征融合 (第一个卷积层精简特征图的通道数量, 去除不必要的信息, 减少后

续处理的计算负担; 第二个卷积层保持特征图的通道数量不变, 维持信息的完整性, 进一步细化特征表达); 最后, 通过 Sigmoid 函数生成空间权重, 并将其应用于原始输入特征图, 以增强空间感知能力。

### 1.3 改进的 DeepLabv3+模型

DeepLabv3+模型以采样距离为 2 的标准卷积对输入特征图进行下采样, 并通过降低输入特征图的尺寸来减少模型计算量。该模型的编码器部分融合了 Xception 架构和空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 单元的优势。Xception 架构用于深入提取图像的深层语义信息; ASPP 单元通过多分辨率特征采样, 保持数据的丰富度。然而, 这种融合方式会牺牲图像的某些细节特征, 尤其是需要精确捕捉的道路分界线。

为了提高模型的分割精度, 并解决 DeepLabv3+模型因高复杂度而导致的训练效率低、缺乏上下文信息聚合的问题, 本文对 DeepLabv3+模型进行以下改进:

1) 考虑到后期需要将模型部署到嵌入式设备端, 受边缘设备资源的限制, 在主干网络部分, 用轻量级的 MobileNetV2 代替 Xception, 减少了模型的参数量和计算需求;

2) 原始 ASPP 单元包含空洞率分别为 6、12 和 18 的 3 个空洞卷积层, 对于大目标分割, 高空洞率分割效果更好; 而对于小目标分割, 低空洞率更有效; 随着主干网络提取的特征图分辨率逐渐降低, 需要低空洞率卷积来提取较低分辨率特征图的信息; 为此, 在 ASPP 单元中新增一条卷积支路, 且空洞率变为 4、8、12 和 16, 提升了模型对多尺寸目标的分割能力;

3) 在解码阶段引入 GAM, 进一步提高模型的分割精度; GAM 使模型捕捉到长距离的依赖关系, 增强对全局上下文信息的聚合能力; 通过 GAM 对特征图不同区域分配不同的权重, 促进不同尺度特征图之间的信息交流, 实现多尺度特征的融合; 特别是在处理具有复杂背景图像时, GAM 能准确地理解图像内容。改进的 DeepLabv3+模型结构如图 4 所示。

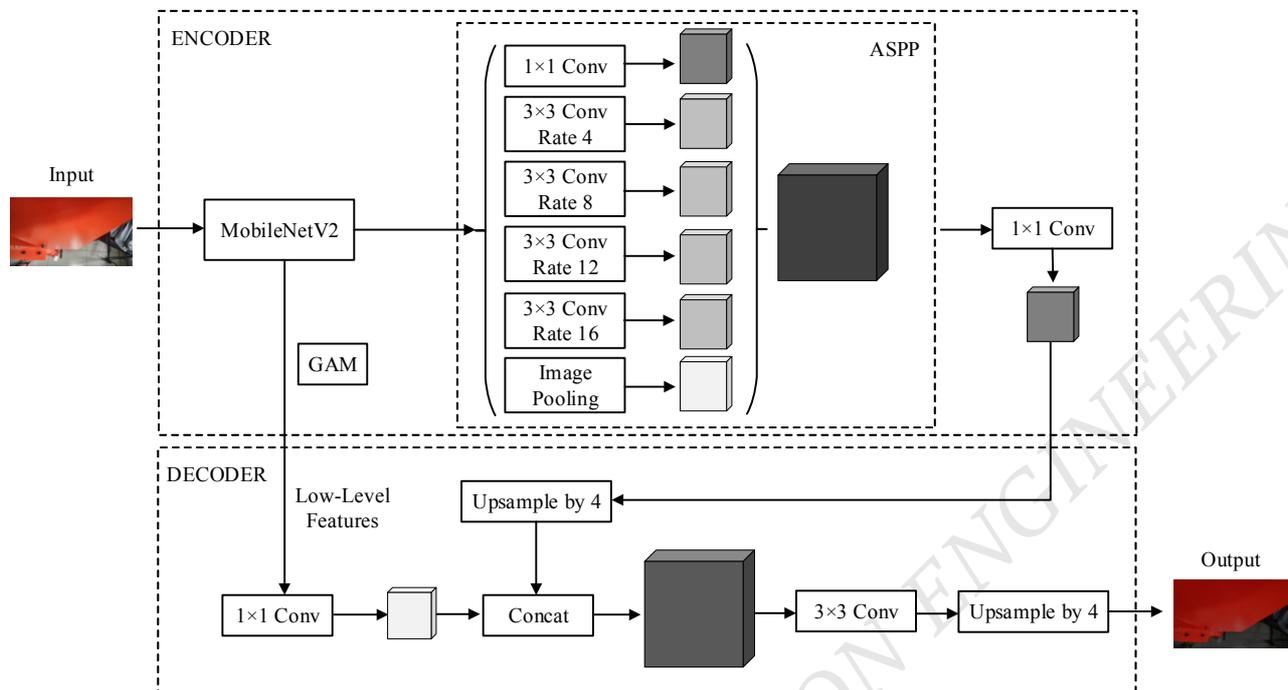


图4 改进的DeepLabv3+模型结构

## 2 实验准备与评价指标

### 2.1 实验准备

本实验采用自制的数据集验证改进的 DeepLabv3+模型的有效性。数据集构建过程如下：

- 1) 攀爬机器人搭载摄像头，在多个不同外形与结构的户外起重机上录制其攀爬过程的视频；
- 2) 利用 Python 对每个视频以固定的帧率截图保存，并将保存的图像作为原始数据集；
- 3) 采用旋转、裁剪、灰度化、滤波等处理方式对原始数据集进行增强，增强后的数据集共有 1 814 幅图像；
- 4) 通过 Labelme 软件对增强后的数据集进行手工标注，可行域被标注为 Safe，背景区域被标注为 Back-ground；
- 5) 将标注后的数据集按照 9 : 1 的比例，随机划分为训练集和测试集。

实验的硬件环境为：CPU 为 E5-2650v4@2.20 GHz，内存为 16 GB，显卡为 NVIDIA Corporation TU102 [TITAN RTX]；软件环境为：Ubuntu20.04，PyTorch1.8.0，Python3.8.5，CUDA 的版本为 11.6。初

始学习率为  $2 \times 10^{-4}$ ，训练批次为 16 个，迭代轮次 epoch 为 200 次，网络优化器采用 Adam。

### 2.2 评价指标

本文采用平均交并比 (MIoU)、平均精确率 (Mpa) 和召回率 (recall) 3 个性能指标来评价攀爬机器人可行域的检测性能。其中，MIoU 为预测结果与实际结果交集与并集的比例关系；Mpa 为各类别像素识别准确度的平均值；recall 为正确预测为可行域占全部实际可行域的比例，计算公式分别为

$$MIoU = \frac{1}{n+1} \sum_{i=0}^n \frac{TP_i}{TP_i + FN_i + FP_i} \quad (2)$$

$$Mpa = \frac{1}{n} \sum_{i=1}^n \frac{TP_i}{TP_i + FN_i + FP_i} \quad (3)$$

$$R_{\text{recall}} = \frac{TP}{TP + FN} \quad (4)$$

式中：TP 为正确预测为可行域的像素总数，FP 为错误将背景区域预测为可行域的像素数量，FN 为将可行域预测为背景区域的像素数量，n 为模型需要识别的目标类别总数。

### 3 实验与分析

为评估本文提出的改进的 DeepLabv3+模型的检测性能,将其与 U-Net 模型、DeepLabv3+模型进行对比实验。在相同的数据集和软硬件环境下运行上述 3 种模型,实验结果如表 2 所示。

模型	$MIoU$	$Mpa$	$R_{recall}$
U-Net	94.28	89.93	94.67
DeepLabv3+	95.82	97.35	96.44
本文模型	<b>97.04</b>	<b>98.50</b>	<b>98.38</b>

由表 2 可知,本文模型比 U-Net 模型的  $Mpa$  和  $R_{recall}$  分别提高了 8.57%和 3.71%;比 DeepLabv3+模型的  $MIoU$  和  $Mpa$  分别提高了 1.22%和 1.15%。

为更直观地展示本文模型(改进的 DeepLabv3+模型)与原始 DeepLabv3+模型的性能差异,将分割结果进行可视化处理,如图 5 所示。

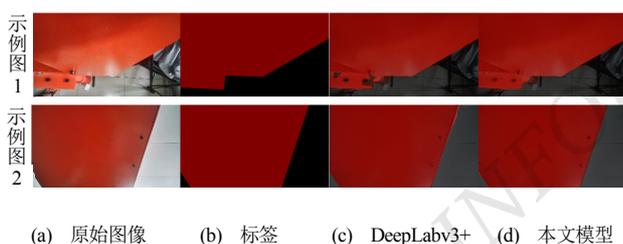


图 5 不同模型的分割结果

由图 5 可以看出:在示例图 1 中,针对螺钉的干扰,DeepLabv3+模型的分割结果不够准确;在示例图 2 中,DeepLabv3+模型虽然基本正确地分割了边界板,但在右侧两块钢板连接处的分割不够完整;本文模型能更准确地捕捉到轮廓,分割结果更清晰,说明本文模型不仅准确地划分了可行域,还聚合了上下文信息,使整个可行域的分割更为完整。

综合实验结果与分割结果图,本文模型与 U-Net 模型、DeepLabv3+模型相比,在可行域分割性能上具有较好的效果。

### 4 结论

本文面向自主移动攀爬机器人在起重机高空作

业的场景,利用摄像头采集实时环境信息,采用改进的 DeepLabv3+模型进行可行域分割,协助攀爬机器人自主移动。实验结果表明,本文模型相较于 U-Net 模型、DeepLabv3+模型,具有更好的分割效果。尽管本文提出的改进的 DeepLabv3+模型提高了可行域分割效果,但其泛化能力仍有待提升。在未来的工作中,考虑提升模型的泛化性能,并将模型部署至嵌入式端,通过 TensorRT 推理框架使模型在机载平台上实时分割可行域,实现攀爬机器人的高效自主移动。

©The author(s) 2024. This is an open access article under the CC BY-NC-ND 4.0 License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

### 参考文献

- [1] 李林.特种设备检测中无损检测技术的运用分析[J].装备制造技术,2024(7):139-141.
- [2] CHEN L, LI S, BAI Q, et al. Review of image classification algorithms based on convolutional neural networks[J]. Remote Sensing, 2021,13(22):4712.
- [3] MINAEE S, BOYKOV Y, PORIKLI F, et al. Image segmentation using deep learning: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021,44(7):3523-3542.
- [4] 杨洁洁,杨顶.基于深度学习的语义分割综述[J].长江信息通信,2022,35(2):69-72.
- [5] ARBCLÁEZ P, HARIHARAN B, GU C, et al. Semantic segmentation using regions and parts[C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition, 2012:3378-3385.
- [6] OLAF R, PHILIPP F, THOMAS B. U-Net: Convolutional Networks for Biomedical Image Segmentation[J]. CoRR, 2015, abs/1505.04597.
- [7] 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.
- [8] LIN G, LIU F, MILAN A, et al. RefineNet: Multi-Path refinement networks for dense prediction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019,42(5):1.

(下转第 46 页)