

# 改进 DeepLabV3+ 网络的指针轨迹图像识别

袁帅<sup>1</sup>, 蒋强<sup>1</sup>, 饶兵<sup>2</sup>

(1. 沈阳理工大学 自动化与电气工程学院, 沈阳 110159; 2. 沈阳天眼智云智能技术研究院有限公司, 沈阳 110179)

**摘要:** 指针式机械记录仪通常用于记录精密设备运输过程中的震动轨迹图像, 为了更好地监测运输过程中车辆颠簸对仪器设备的影响, 提出一种改进 DeepLabV3+ 网络的指针轨迹图像语义分割方法。首先将骨干网络替换为 MobileNetV3, 实现模型的轻量化; 然后将解码器中 4 倍上采样替换为 2 次 2 倍上采样, 增强图像中像素的连续性, 使预测结果更接近原始图像。在自制数据集上进行对比实验, 结果表明: 改进 DeepLabV3+ 网络的平均交并比 (MIoU) 达到 85.84%, 比原始 DeepLabV3+ 网络提高了 3.57%, 单位时间内检测图片数量 (FPS) 提高了  $3.58\text{ s}^{-1}$ ; 改进 DeepLabV3+ 网络在识别精度和速度上具有明显的优势, 可为精密仪器检测提供数据支持。

**关键词:** 改进 DeepLabV3+; 语义分割; 轨迹图像识别; 轻量化

**中图分类号:** TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1003-1251.2024.01.008

## Pointer Trajectory Recognition with Improved DeepLabV3+ Network

YUAN Shuai<sup>1</sup>, JIANG Qiang<sup>1</sup>, RAO Bing<sup>2</sup>

(1. Shenyang Ligong University, Shenyang 110159, China; 2. Shenyang SkyEye Intelligent Cloud Technology Research Institute Co., Ltd., Shenyang 110179, China)

**Abstract:** Pointer-type mechanical recorders are usually used to record images of vibration trajectories during the transportation of precision equipment. In order to better monitor the impact of vehicle bumps on instruments and equipment during transportation, a semantic segmentation method of pointer trajectory images with improved DeepLabV3+ network is proposed. Firstly, the backbone network is replaced with MobileNetV3 to realize the lightweight of the model. Then the 4-fold up-sampling in the decoder is replaced with 2 times 2-fold up-sampling to enhance the continuity of pixels in the image, which makes the predicted results closer to the original image. The results show that the average intersection ratio (MIoU) of the improved DeepLabV3+ network reaches 85.84%, which is 3.57% higher than that of the original DeepLabV3+ network, and the number of detected images per unit time (FPS) increases by  $3.58\text{ s}^{-1}$ . The improved DeepLabV3+ network has obvious advantages in recognition accuracy and speed, which can provide data support for precision instrument detection.

**Key words:** improved DeepLabV3+; semantic segmentation; trajectory image recognition; lightweight

在精密仪器运输过程中,通常采用指针式机械记录仪记录震动轨迹图像数据,若采用人工检定方法,检定结果不精准,工作量巨大且效率很低。通过机器学习相关技术,可实现高效率高精度自动检测指针式机械记录仪震动轨迹图像,极大地减少人工工作量<sup>[1]</sup>。

2014 年开始,谷歌团队推出并发展了具有良好分割性能的 DeepLab 系列分割网络,其中 DeepLabV3 + 网络<sup>[2]</sup>在语义分割领域表现更为突出,准确度更高。以 DeepLabV3 + 为基础,学者们进行了深入研究,将其应用于诸多领域。2019 年, Liu 等<sup>[3]</sup>通过添加更多的跳跃连接和卷积层来设计 DeepLabV3 + 解码器,改善了遥感图像中建筑物轮廓的检测结果,但对细微边界分割效果不理想。2022 年, Zhang 等<sup>[4]</sup>在 DeepLabV3 + 中加入一种基于边缘信息的损失函数,提高了网络对舌边的分离效果,但对错误分类处理能力不高。同年,刘慧等<sup>[5]</sup>使用轻量化 MobileNetV2 作为 DeepLabV3 + 骨干网络,减少了模型参数,并使用 ReLU6 激活函数减少部署在移动端设备上的精度损失,但对小像素目标识别效果较差。2023 年,周迅等<sup>[6]</sup>在 DeepLabV3 + 中使用三点注意力模块提高了对坝面裂缝像素的提取能力,但存在漏检情况。

本文将 DeepLabV3 + 网络模型应用于指针式机械记录仪轨迹图像的识别。使用轻量化网络 MobileNetV3 替换原骨干网络 Xception,实现模型轻量化;在解码器中使用 2 个连续的 2 倍上采样替换原网络中的 4 倍上采样,将提取到的特征图逐步放大,使得还原出的边界更加细化。

## 1 DeepLabV3 + 网络及其改进

### 1.1 DeepLabV3 + 网络介绍

DeepLabV3 + 网络使用编码器 - 解码器 (Encoder-Decoder) 结构<sup>[7-8]</sup>,在提升分割效果的同时关注边界信息。模型采用 Xception 作为骨干网络,使用空洞空间卷积金字塔池化 (atrous spatial pyramid pooling, ASPP) 融合特征图多尺度信息,并将深度卷积和逐点卷积<sup>[9]</sup>应用于 ASPP 和 Encoder 模块中,使网络训练速度更快。DeepLabV3 + 网络结构如图 1 所示。图中: Conv 表示卷积;rate 表示膨胀率;Upsample 表示上采样;DCNN 表示深度卷积神经网络;Atrous Conv 表示空洞卷积;Pooling 表示池化;Low-level Fea-

ture 表示低级特征;Concat 表示数据拼接。

DeepLabV3 + 网络通过 Encoder 结构得到两部分图片特征,在 Decoder 中使用卷积调整通道,融合两部分特征,再使用线性插值上采样使得输出层和原图片尺寸一致,获得预测结果<sup>[10-11]</sup>。

### 1.2 改进的 DeeplabV3 + 网络

#### 1.2.1 改进 DeepLabV3 + 骨干网络

由于本文检测对象是设备运输过程中的指针震动轨迹,为满足实时检测和移动检测的要求,需将训练好的模型部署在移动端硬件平台上。因此,在进行图像特征提取时要尽量保证全局信息的准确性,同时简化参数和计算量,保证识别效率。

为解决上述问题,可采用轻量化网络模型。DeepLabV3 + 骨干网络为 Xception,该网络结构比较复杂,参数量较多,消耗大量的显存。本文对 DeepLabV3 + 结构的骨干网络实现轻量化,采用 MobileNetV3 代替 Xception<sup>[12]</sup>,轻量化网络 MobileNetV3 的参数量较少,更易于部署到移动设备上。MobilenetV3 的瓶颈 (bneck) 结构如图 2 所示。图中:NL 表示非线性激活函数;Pool 表示平均池化;Dwise 表示深度可分离卷积;FC 表示全连接层;ReLU、hard- $\sigma$  表示激活函数; $\otimes$  表示乘法操作。

MobilNetV3 在 MobilNetV2 的结构基础上增加了注意力机制 (squeeze-and-excitation, SE),并使用 h-swish 激活函数替换 swish 函数,相比于 MobilNetV2, MobilNetV3 模型更加轻量化,降低了计算成本,识别精度更高,计算速度更快。

#### 1.2.2 改进 DeepLabV3 + 模型解码器设计

经 Encoder 得到的特征图由大量的像素矩阵构成,各像素之间均存在密切的联系,DeepLabV3 + 网络中 Decoder 将传入的特征图直接使用一次双线性 4 倍上采样恢复目标边界信息,会使图像的像素不连续,导致网络预测边界不精确<sup>[13]</sup>。本文将传入 Decoder 中的特征信息先进行一次 2 倍上采样,还原边界信息,然后再进行一次 2 倍上采样,即使用 2 个连续的 2 倍上采样替换原始 DeepLabV3 + 网络中的 4 倍上采样,增强图像中像素的连续性,还原出的边界信息更接近原始标注图像,从而获得更清晰的目标边界。改进的 DeepLabV3 + 网络模型在 Decoder 中仅添加了 1 次上采样操作,相比 DeepLabV3 + 网络模型,参数增加极少。改进前后的 Decoder 部分如图 3 所示。

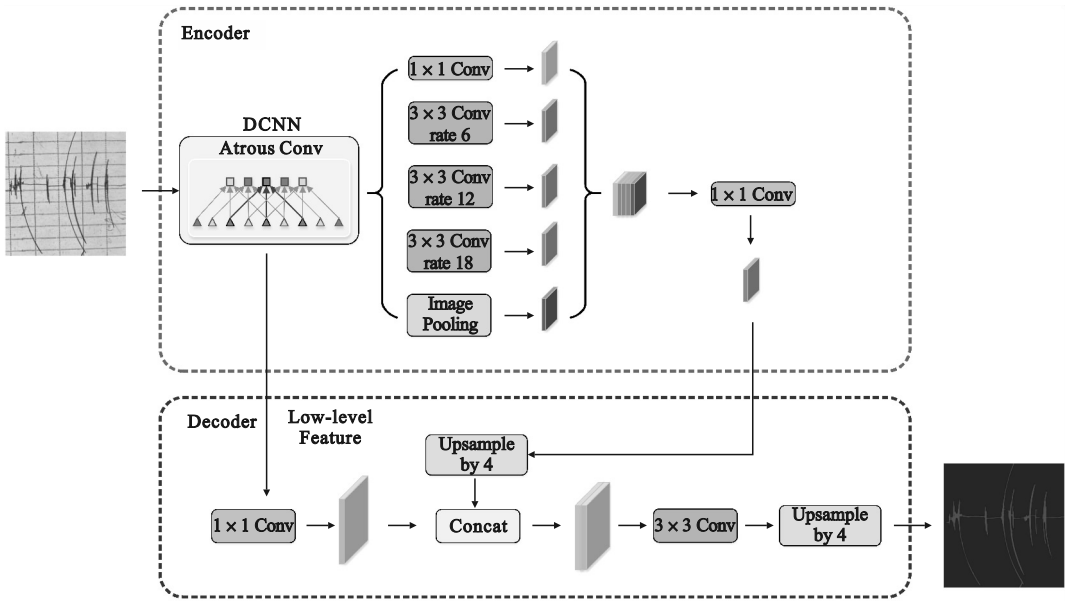


图1 DeepLabV3+ 网络结构

Fig. 1 DeepLabV3+ network structure

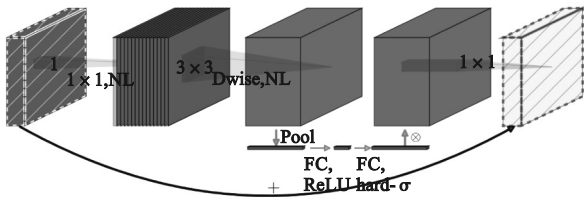


图2 MobileNetV3 网络的 bneck 结构

Fig. 2 Bneck structure of MobileNetV3 network

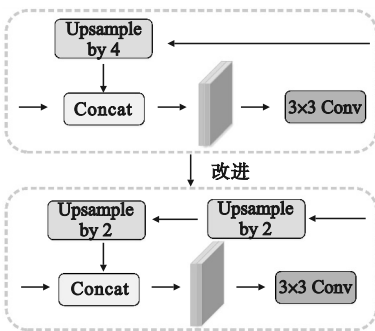


图3 改进前后的 Decoder 部分

Fig. 3 Decoder part before and after improvement

改进后的 DeepLabV3+ 网络结构如图 4 所示。

## 2 实验与分析

### 2.1 实验环境

本文基于百度飞桨 (PaddlePaddle) 深度学习

框架进行实验。模型训练过程中使用随机梯度下降法,最大训练轮次为 10 000 轮。单次训练样本数设置为 2,初始学习率设置为 0.01,之后通过多项式衰减策略减少学习率。

实验数据来自设备运输过程中采集的指针式机械记录仪震动轨迹图像,以此自制数据集,该数据集包含精细标注影像 873 张,标注内容包括轨迹和背景。由于总体样本数据较少,采用图像反转、水平和垂直镜像等处理方法进行数据增强,丰富实验数据集。

为解决网络模型在不同工作场景的算法适用性问题,使用图像分割套件 PaddleSeg 中的预训练模型,加快模型训练速度并保证特征提取效果,提高模型对指针轨迹图像识别的准确性和泛化性。

### 2.2 评价指标

本文选用平均交并比 (MIoU) 作为实验结果评价指标,MIoU 是语义分割领域的标准度量指标。分别对每个类计算交并比 (IoU),再对所有类别的 IoU 求均值,即为 MIoU。其计算式为

$$MIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - P_{ii}} \quad (1)$$

式中: $k+1$  表示图像中所有分割类别数目; $p_{ij}$  表示标签为  $i$  被预测为  $j$  的样本数量; $p_{ji}$  表示标签为  $j$  被预测为  $i$  的样本数量; $p_{ii}$  表示标签为  $i$  被预测为  $i$  的样本数量。

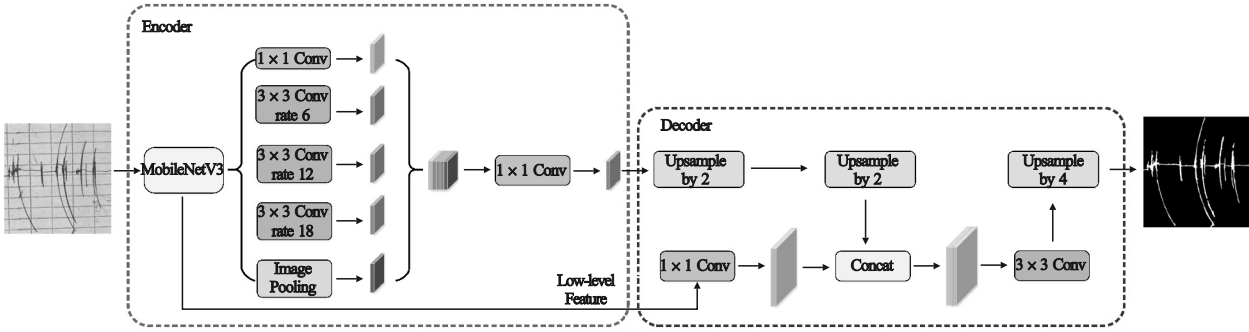


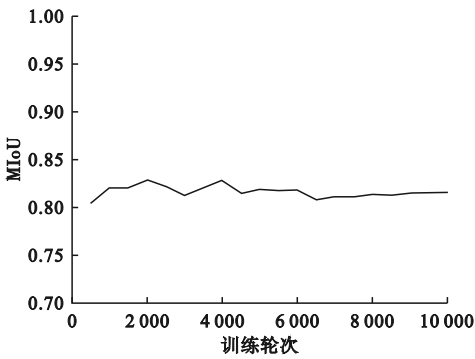
图 4 改进后的 DeepLabV3+ 网络结构

Fig. 4 Improved DeepLabV3+ Network Structure

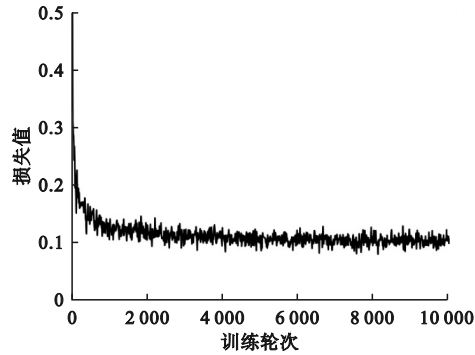
采用单位时间内模型检测图片的数量(FPS)作为另一个评价指标,该评价指标越大,表示检测的速度越快,其值为待检测的图片总数与模型预测所需的时间之比。

2.3 实验对比

对改进前后的 DeepLabV3+ 模型进行训练测试,两组实验均使用 PaddleSeg 套件中自带的预训练模型,训练过程中其他条件相同,得到的可视化参数如图 5 和图 6 所示。



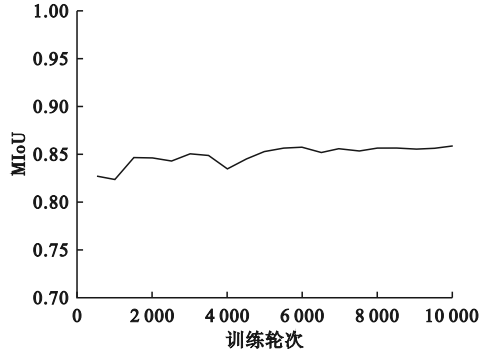
(a) 平均交并比



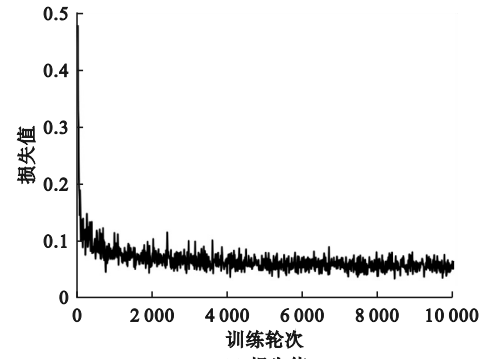
(b) 损失值

图 5 原始 DeepLabV3+ 网络训练参数  
Fig. 5 Original DeepLabV3+ network training parameters

由图 5 和图 6 可见,改进 DeepLabV3+ 网络的 MIoU 更高,模型识别指针轨迹图像的能力更好,损失函数能够更快地收敛且下降程度更大,模



(a) 平均交并比



(b) 损失值

图 6 改进 DeepLabV3+ 网络训练参数  
Fig. 6 Improved DeepLabV3+ network training parameters

型的性能更优。

训练完成后,将测试集中图片统一调整为相同的分辨率,分别使用网络改进前后两组实验中的最好模型进行预测,以图片 1 和图片 2 为例,两个图片的预测结果如图 7 和图 8 所示。

由图 7 和图 8 可见,本文提出的改进 DeepLabV3+ 网络模型可以更好地获取轨迹语义信息,漏检和误检的比例更小,对轨迹特征提取能力更强,能够更好地完成轨迹图像识别任务。

模型预测参数及模型大小如表 1 所示。

由表 1 可知,本文提出的改进 DeepLabV3+ 网络模型检测结果更精细、检测速度更快、模型体

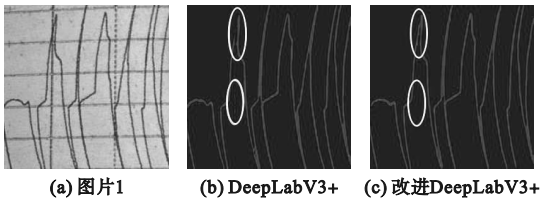


图7 对比实验预测结果(图片1)

Fig. 7 Comparisons of experimental prediction results (image 1)

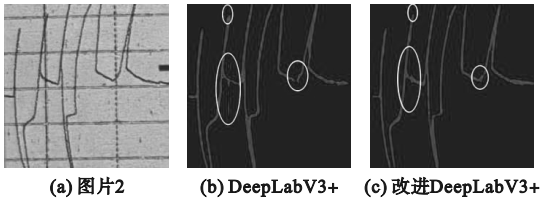


图8 对比实验预测结果(图片2)

Fig. 8 Comparisons of experimental prediction results (image 2)

表1 模型参数对比

Table 1 Comparisons of model parameters

| 模型             | MIOU/% | FPS/s <sup>-1</sup> | 模型大小/MB |
|----------------|--------|---------------------|---------|
| DeepLabV3 +    | 82.88  | 2.73                | 176.91  |
| 改进 DeepLabV3 + | 85.84  | 6.31                | 47.27   |

积更小,体现了该算法的可行性和优越性。

### 3 结论

为解决运输过程中车辆颠簸是否导致仪器设备损坏的检测问题,以 DeepLabV3 + 作为语义分割模型,将其骨干网络替换为轻量化网络 MobileNetV3,减少了参数量,解决了模型部署到移动端硬件平台的问题;针对 Decoder 结构中 4 倍上采样操作导致图像中的边界像素不连续、丢失某些重要像素信息问题,采用 2 次 2 倍上采样增强图像中像素的连续性,还原出的边界信息更接近原始标注图像,获得了更清晰的目标边界。改进后的 DeepLabV3 + 模型体积减少了 129.64 MB;检测结果更精细,MIoU 达到 85.84%,提高了 3.57%;检测速度更快,FPS 提升了 3.58 s<sup>-1</sup>。本文提出的改进 DeepLabV3 + 网络降低了模型的参数量、加快了检测速度、提高了模型对轨迹图像识别的泛化能力,更宜于实际应用。

#### 参考文献 (References):

[1] 王宇. 运动轨迹检测识别技术研究[D]. 沈阳:东北大学,

2015.

- [2] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation[C]//Proceedings of the European Conference on Computer Vision. Munich, Germany: Springer, 2018: 801-818.
- [3] LIU J Q, WANG Z L, CHENG K X. An improved algorithm for semantic segmentation of remote sensing images based on DeepLabv3+[C]//Proceedings of the 5th International Conference on Communication and Information Processing. New York, USA: ACM, 2019: 124-128.
- [4] ZHANG X F, BIAN H N, CAI Y H, et al. An improved tongue image segmentation algorithm based on Deeplabv3 + framework[J]. IET Image Processing, 2022, 16(5): 1473-1485.
- [5] 刘慧,姜建滨,沈跃,等.基于改进 DeepLab V3 + 的果园场景多类别分割方法[J].农业机械学报,2022,53(11): 255-261.
- LIU H, JIANG J B, SHEN Y, et al. Multi-category segmentation of orchard scene based on improved DeepLab V3+[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(11): 255-261. (in Chinese)
- [6] 周迅,李永龙,周颖玥,等.基于改进 DeepLabV3 + 网络的坝面裂缝检测方法[J/OL].清华大学学报(自然科学版), 2023; 1-11 [2023-04-07]. https://kns.cnki.net/kcms/detail//11.2223.N.20230208.1512.001.html. DOI: 10.16511/j.cnki.qhdxxb.2023.26.006.
- ZHOU X, LI Y L, ZHOU Y Y, et al. Dam surface crack detection method based on improved DeepLab-V3 + network[J/OL]. Journal of Tsinghua University (Science and Technology), 2023; 1-11 [2023-04-07]. https://kns.cnki.net/kcms/detail//11.2223.N.20230208.1512.001.html. DOI: 10.16511/j.cnki.qhdxxb.2023.26.006. (in Chinese)
- [7] 周倩倩,刘汉林,陈维锋,等.基于 Deeplabv3 + 的排水管道缺陷检测与语义分割[J].中国给水排水,2022,38(13): 22-27.
- ZHOU Q Q, LIU H L, CHEN W F, et al. Drainage pipeline defects detection and semantic segmentation based on DeepLabv3+[J]. China Water & Wastewater, 2022, 38(13): 22-27. (in Chinese)
- [8] 赵祥,王涛,张艳,等.基于改进 DeepLabv3 + 孪生网络的遥感影像变化检测方法[J].地球信息科学学报,2022,24(8):1604-1616.
- ZHAO X, WANG T, ZHANG Y, et al. Remote sensing image change detection based on improved DeepLabv3 + Siamese network[J]. Journal of Geo-information Science, 2022, 24(8): 1604-1616. (in Chinese)
- [9] 马冬梅,李鹏辉,黄欣悦,等.改进 DeepLabV3 + 的高效语义分割[J].计算机工程与科学,2022,44(4):737-745.
- MA D M, LI P H, HUANG X Y, et al. Efficient semantic segmentation based on improved DeepLabV3+[J]. Computer Engineering and Science, 2022, 44(4): 737-745. (in Chinese)
- [10] 姚燕,胡立坤,郭军.基于改进 DeepLabv3 + 网络的轻量级语义分割算法[J].激光与光电子学进展,2022,59(4):200-207.
- YAO Y, HU L K, GUO J. Improved lightweight semantic segmentation algorithm based on DeepLabv3 + network[J]. Laser & Optoelectronics Progress, 2022, 59(4): 200-207. (in Chinese)

(下转第60页)