

基于机器视觉的台架上钢坯位置分割^①



张 哲¹, 邵允学¹, 吕 刚²

¹(南京工业大学 计算机科学与技术学院, 南京 211816)

²(上海策立信息技术有限公司, 上海 201900)

通信作者: 邵允学, E-mail: csshxyx@njtech.edu.cn

摘 要: 钢坯通过航车从库存调度到台架, 然后从台架经轨道到达炉前, 以往是人工控制机械将台架上的钢坯推到轨道上的. 这个过程的自动化实现需要知道钢坯在台架上的实时的位置分布, 以便于自动控制推钢机. 本文通过机器视觉方法实现台架上钢坯的实时定位, 提出了以 U-Net 为基础网络, 结合经典 ResNet 网络中的残差块, 实现了钢坯横向位置的精确分割. 实验结果和现场应用测试表明, 本文方法的分割精度能够达到工业现场的控制需求.

关键词: 深度学习; 残差网络; Res-UNet; 目标分割; 钢坯定位

引用格式: 张哲, 邵允学, 吕刚. 基于机器视觉的台架上钢坯位置分割. 计算机系统应用, 2022, 31(10): 254-260. <http://www.c-s-a.org.cn/1003-3254/8782.html>

Position Segmentation of Billet on Bench Based on Machine Vision

ZHANG Zhe¹, SHAO Yun-Xue¹, LYU Gang²

¹(School of Computer Science and Technology, Nanjing Tech University, Nanjing 211816, China)

²(Shanghai Celi Engineering Technology Co. Ltd., Shanghai 201900, China)

Abstract: A billet is dispatched from the inventory to the bench by a crane and then from the bench to the front of the furnace through a track. In the past, the billet was pushed onto the track by the manual control of machinery. The automation of this process requires knowledge of the real-time position distribution of billets on the bench for automatic control of the pusher. In this study, the real-time positioning of billets on the bench is achieved by the machine vision method. Specifically, with the U-Net as the basic network, the residual blocks in classic ResNet are used to achieve the accurate segmentation of transverse positions of billets. The experimental results and field application tests indicate that the segmentation accuracy of this method can meet the control requirements of industrial fields.

Key words: deep learning; residual network; Res-UNet; object segmentation; billet positioning

钢铁作为国内的支柱产业, 对于我国的经济发展具有推动作用, 在建筑、航空航天、交通、桥梁等多个领域都要有着重要的角色地位. 然而, 国内大部分的钢铁生产流程中还存在人工劳动强度大、成本高, 生产效率低, 自动化和智能化程度不足等问题. 以台架上钢坯被推到轨道这个过程为例, 当前生产流程需要依靠人眼观察钢坯在台架上的位置分布, 然后手动控制推钢机到达指定位置, 然后手动控制将对应钢坯推向轨道. 这个流程人工工作强度大、易疲劳, 生产工作效率较低.

计算机视觉作为人工智能领域的重要组成部分, 主要作为模仿人类视觉功能的工具, 通过对目标图像进行特征提取并加以分析理解, 最终实现对目标的分类识别. 目前计算机视觉是一个广泛而复杂的研究领域, 其技术已经应用于众多领域的研究以及生活生产环境中. 计算机视觉主要的任务有图像分类^[1]、目标分割^[2]、图像定位^[3,4]、目标检测^[5,6]、语义分割^[7]等. 目标分割作为本文的研究课题, 传统的方法包括阈值法^[8,9]、边界检测法^[10]、区域法^[11]等. 近些年随着深度

① 收稿时间: 2022-01-14; 修改时间: 2022-03-21; 采用时间: 2022-04-02; csa 在线出版时间: 2022-07-07

学习算法的发展^[12],目标分割等机器视觉任务都有了长足的进步。

本文的主要创新点在于对 ResNet 网络中的残差块进行改进;并采用 U-Net 网络的设计思路,把整个网络分成特征提取部分与上采样两个部分.本文搭建的网络模型充分利用了两者的优点,相关内容将在以下章节进行详细介绍.

1 相关工作

深度学习的发展为计算机视觉这一领域提供了诸多解决方案,带来鲁棒性、精确度更高的模型算法.作为深度学习的主要模型,卷积神经网络(CNN)由于其出色的特征提取能力,近些年流行用于计算机视觉领域.卷积神经网络相较于常规的神经网络模型,卷积神经网络可以在计算机视觉领域中取得成功有两个重要的原因:首先它利用了2D图像结构和在邻域内的像素具有高度相关性的性质,此外它的体系结构依赖于特征共享,因此每个通道(或输出特征映射)都是通过卷积在所有位置使用相同大小的滤波器生成的;其次是它引入了池化层的操作,提供了一定程度的平移不变性,减少了因位置变化给模型带来的影响,池化操作还可以扩大网络的感受野,使得网络可以看到更大部分的输入.

AlexNet^[13]由Krizhevsky等人设计,作为2012年ILSVRC竞赛的冠军网络,取得top-5错误率为15.3%的成绩.AlexNet的成功使得计算机视觉领域的研究人员开始广泛采用卷积神经网络模型并不断对其结构进行改进,取得了巨大的突破.AlexNet的成功在于该网络首次利用GPU进行网络加速训练;使用了ReLU激活函数,代替了传统的Sigmoid激活函数以及tanh激活函数;使用了LRN局部响应归一化;在全连接层的前两层中使用了Dropout随机失活神经元操作,以减少过拟合.AlexNet该模型一共分为8层,5个卷积层,以及3个全连接层,在每一个卷积层中包含了ReLU激活函数以及局部响应归一化(LRN)处理,然后再经过降采样(池化处理)降低输出特征图的大小.

VGGNet^[14]在2014年由牛津大学著名研究组VGG(visual geometry group)提出,在ILSVRC竞赛上取得top-5错误率为6.8%的成绩.VGGNet网络最大的特点设计通过组合和堆叠 3×3 卷积核滤波器,对输入进行高效的特征提取.相较卷积核大小为 5×5 和 7×7 的卷积,VGG通过组合多个 3×3 卷积核实现了相

同的感受野,并且连续小卷积核的卷积比单个大的卷积核的卷积具有更好的非线性.常见的VGG网络有VGG-13、VGG-16、VGG-19,网络层数分别为13层、16层、19层.

1.1 ResNet

ResNet^[15]在2015年由微软研究院的He等人提出,斩获当年ILSVRC竞赛分类任务第一名和目标检测第一名.ResNet首次在卷积神经网络中引入了残差连接,解决了随着网络加深而网络性能却越来越差的问题,构造了一个超过100层的卷积神经网络模型.提出的残差块结构成功地解决了因网络层数的简单堆叠导致的梯度消失、梯度爆炸、退化问题等一系列问题.

残差块结构如图1所示.其中 x 表示残差块的输入, $F(x)$ 表示残差块中隐层输出,ReLU表示激活函数, $F(x)+x$ 表示残差单元输出.残差单元内部连接Batch Norm层对数据进行规范处理,此层实现了权重参数的归一化.

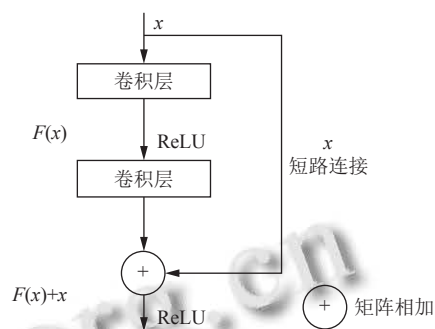


图1 残差块结构

1.2 U-Net

U-Net^[16]提出的初衷是为了解决医学图像分割的问题,解决细胞层面的分割的任务,在2015年的ISBI cell tracking比赛中获得了多项第一.U-Net是基于FCN^[17]改进而得的一种全卷积网络,结构类似于U型,因此得名U-Net.相较于FCN,U-Net在采用上采样方法的基础上,使用对称结构去搭建网络框架.

如图2所示^[16],U-Net网络由左侧的特征提取部分与右侧的上采样部分组成,这种结构称做编码器-解码器结构.其中,每个蓝色矩形代表一个多通道的特征图;蓝色矩形顶部的数字代表的是通道数,左侧数字是指特征图的长度和宽度;白色矩形代表从特征提取部分复制过来的特征图;不同颜色的箭头代表不同的操作.特征提取部分是一个下采样的过程,使用最大池化层

来操作;上采样部分则是一个上采样的过程,采用 2×2 的上卷积(Up-Conv),每次上卷积的输出通道都是原先的一半,再与特征提取部分中对应的特征图拼接

在一起使得通道数与原先大小相同. U-Net 由于其出色的性能和适中的模型规模,其逐渐成为最广泛使用的分割网络并不断进行着改进.

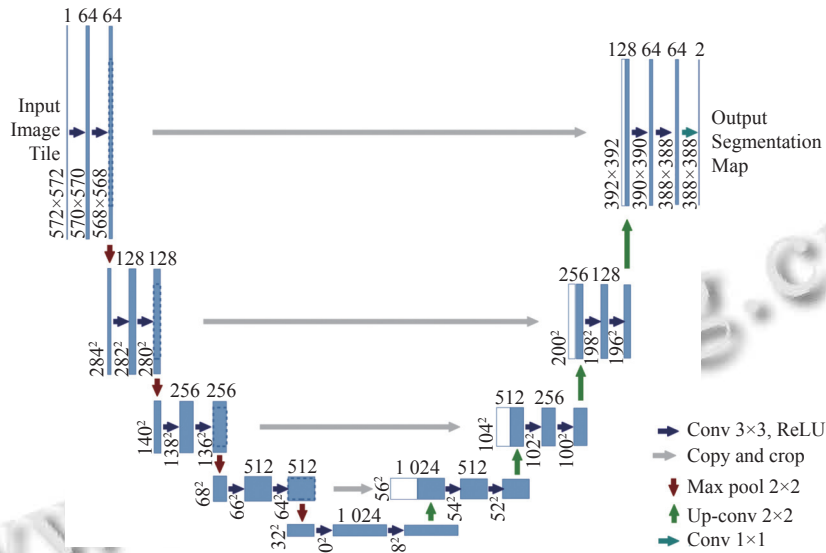


图2 U-Net 网络模型

2 本文方法

2.1 问题描述

钢坯在台架上的移动定位需要去确认钢坯的当前位置以及钢坯数量. 现场采集画面如图3所示,钢坯呈等宽的长条状,且相较垂直于移动方向的角度偏移较小,故可简化成用在图像的移动方向的中线上去表示分割结果.



图3 现场采集画面

2.2 分割模型

本文采用 U-Net 中编码器-解码器的思想,以 ResNet 为基础使用残差连接构造了基础模块,构建了本文的 Res-UNet. 其中编码器的基础块设计沿用了 ResNet 中的设计:使用了卷积核大小为3的卷积层,每个卷积层后进行 BatchNorm 和 ReLU,块与块之间使用残差连接. 解码器的基础块中使用双线性插值的方法对输入特征图进行上采样的操作,这一操作使得特征图的空间维度扩大一倍,其余部分由卷积核大小为3的卷积层组成并使用 BatchNorm 和 ReLU. 整个网络结构类似于 U-Net 的设计,分为特征提取部分和上采样部分,考虑到本文的输入图片尺寸较大,本文在特征提取部

实现思路如图4所示. 本方法将原始图像输入设计的目标分割网络进行运算,输出预测向量. 预测向量表示中线各像素为钢坯的预测值,高于阈值即为钢坯. 线端中黑色线条表示向量中为钢坯的区间,灰色线条表示为其他物体的区间. 之后将向量转换成目标坐标,其中, x_{il} 和 x_{ir} 表示中线处钢坯分割结果的端点值, i 表示钢坯数量, l 和 r 分别代表钢坯的左端点和右端点.

分中使用了更多的基础模块来实现下采样. 本文的改进残差块如图5所示.

本文模型的网络结构如图6所示,由前端的特征提取部分以及后端的上采样部分组成. 特征提取部分由7组本文的改进残差块并联组成,每次的输出矩阵的尺寸宽度和高度为输入的一半,通道数增加一倍. 上采样部分由5组复制拼接操作和改进残差块交替衔接组成,最后连接的输出块输出最终的分割向量. 上采样部分的复制拼接操作是对前一个残差块的输出矩阵以及特征提取部分同层的输出矩阵进行复制拼接操作,此操作输出矩阵作为下次上采样部分残差块的输入矩阵. 上采样部分的残差块使用和特征提取部分相同的残差块.

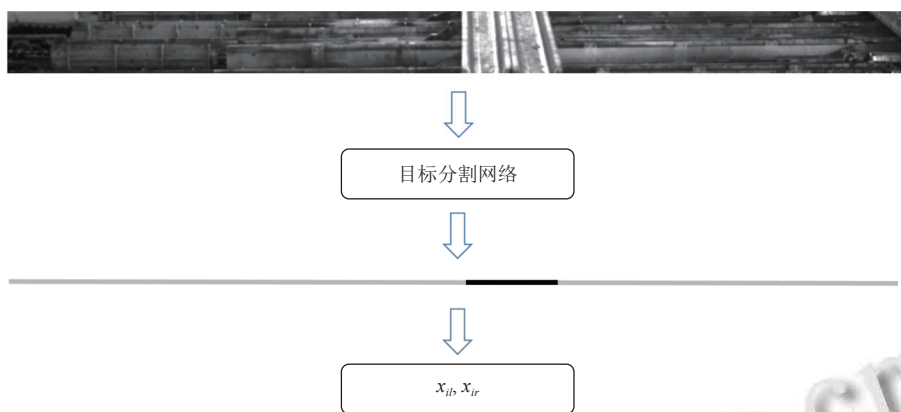


图4 实现思路示意图

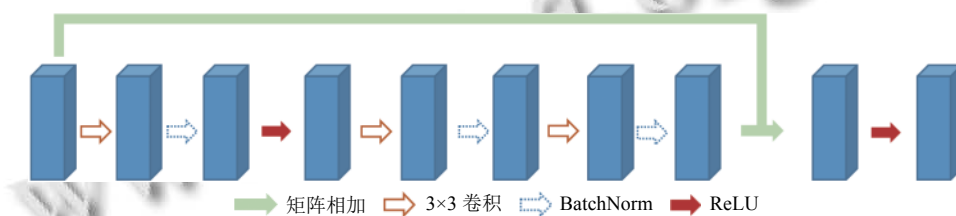


图5 改进残差块

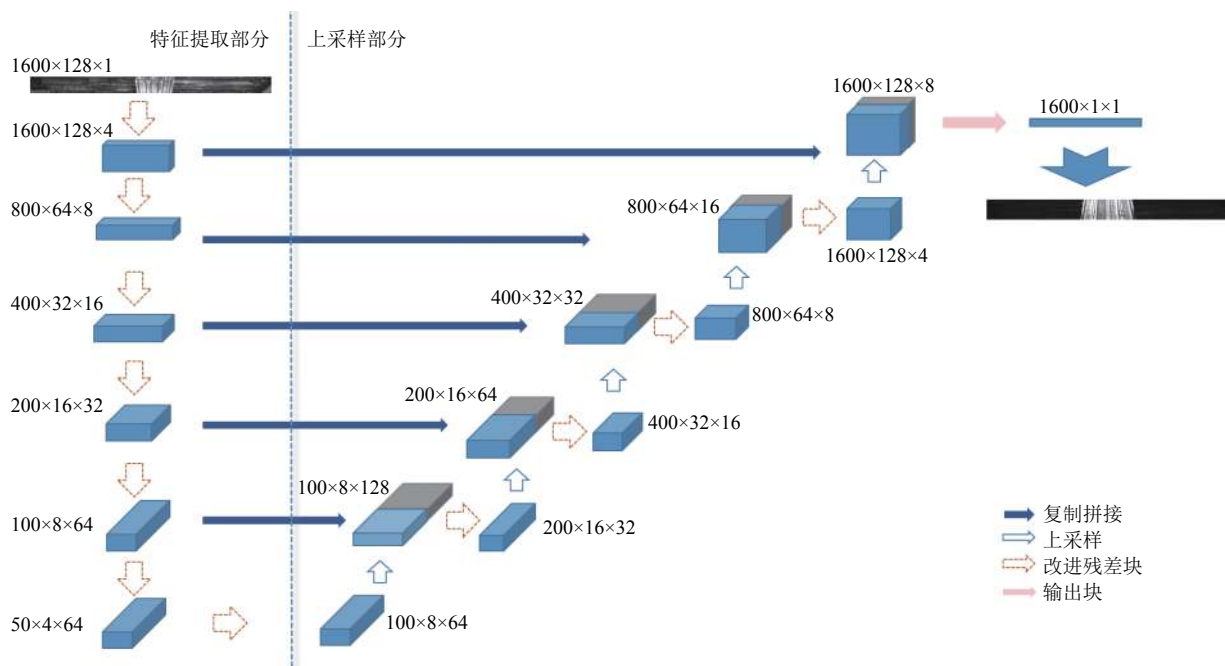


图6 本文模型网络结构

此模型最后的网络模块为输出块, 网络结构如图7所示. 本文设计的输出块使用全局平均池化层以及之前的若干层卷积层共同替代了全连接层, 有效地减少了参数量, 最终输出矩阵尺寸为 $1600 \times 1 \times 1$. 矩阵

第一维的参数表示对应图像中线像素位置的预测钢坯的预测值. 之后, 此预测向量按照阈值量化成用 1 和 0 表示的向量, 若为 1 认定此处像素为钢坯, 为 0 为其他物体.

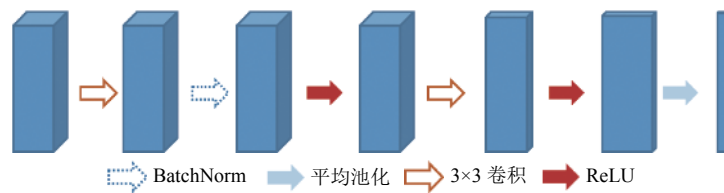


图7 输出块网络结构

3 实验设置与结果分析

3.1 数据集

本文所采用的数据集是从某钢厂真实生产环境中收集的. 在搭建好相机和光源后, 本文通过样本采集程序 1 s 采集 1 张现场的照片, 总共采集 1 周时间, 然后对保存的照片进行筛选, 最后总共得到 48 125 张有效图像样本, 将其中的 45 050 张用来作为训练样本, 3 075 张用来作为测试样本.

在样本标定过程中, 本文采用自己开发的标定软件对所有样本进行标定. 标定软件如图 8 所示, 其中设计亮暗调节按钮调节标定样本的亮度, 方便标注人员对亮度不均匀的样本进行标定. 图中线条为图像垂直方向中线, 即本任务方法的基准线. 线段中的深色部分表示此位置有钢坯, 白色部分表示没有钢坯, 当有多段钢坯时以不同灰度线段交替排列, 方便标注人员观察. 标定结果为每小段钢坯会保存为与图像同名的 TXT 文本文件, 标定输出为每小段钢坯的中线位置归一化后的坐标值.

3.2 实验环境及训练

计算平台搭载 Intel Xeon Gold 5118 处理器, 64 GB 内存, NVIDIA GeForce RTX 2080 Ti 11 GB 显卡 (CUDA

版本为 11.3); 深度学习框架采用 PyTorch 1.6.0. 设计的训练策略使用均方损失函数 (MSELoss) 作为损失函数, 使用 AdaGrad 优化算法作为优化器, 设置初始的学习率为 0.001, 训练迭代次数 epoch 设为 100. 根据计算平台硬件环境, 训练策略设置 batch size 为 32. 主要参数如表 1 所示.

3.3 评价指标

本文采用像素准确率 (pixel accuracy, PA) 作为评价指标, 衡量本文的目标分割模型性能.

像素准确率指在某一类别中, 像素点中正确分割的像素数量和图像像素总量的比例, 如式 (1) 所示:

$$PA = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \quad (1)$$

其中, p_{ij} 表示将第 i 类分成第 j 类的像素数量 (所有像素数量). 像素准确率用于衡量模型在某一类别上的目标分割效果, 数值越大, 模型在此类别上的分割能力越强.

3.4 结果分析

Res-UNet 模型训练耗时约 25 h, 总参数为 300 836 个, 最终的识别定位准确率达 99.7%.

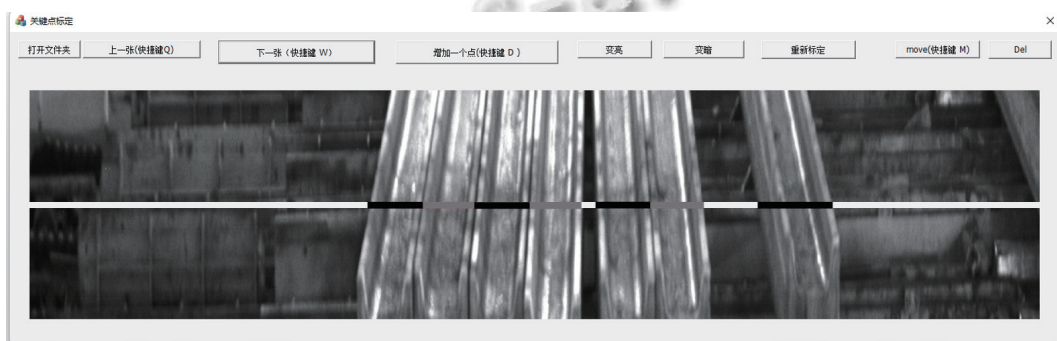


图8 数据样本及标注软件

训练如图 9 所示, 在训练几个周期后, 损失函数的值收敛速度很快, 从 0.010 直接降为 0.002 左右. 通过继续学习, 损失函数的值继续下降, 并稳定在 0.001 左右, 并最终收敛. 测试集训练损失函数最后降到 0.0025

左右且不断震荡, 通过实验发现在大约 30 个周期达到收敛, 最终的准确率约为 99.75%.

3.5 经典 U-Net 比较结果

为了测试所用网络模型的性能, 将本文 Res-UNet

模型与经典 U-Net 模型进行对比实验. 采用相同的训练策下, 经典 U-Net 模型训练准确率如图 10 所示, 最终在测试集上的识别定位准确率达 99.725%. 本文所用网络模型和经典 U-Net 模型在网络参数量、训练耗时以及评价指标的对比如表 2 所示.

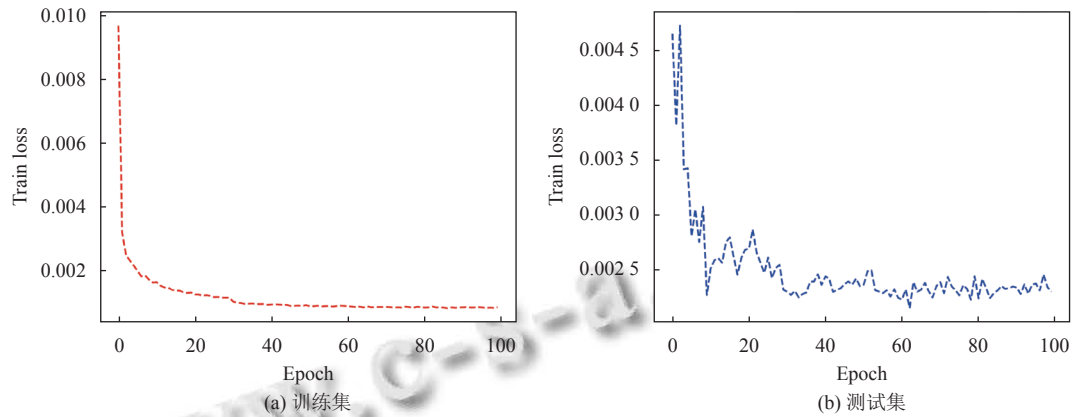


图 9 Res-UNet 模型在训练集和测试集上的训练结果

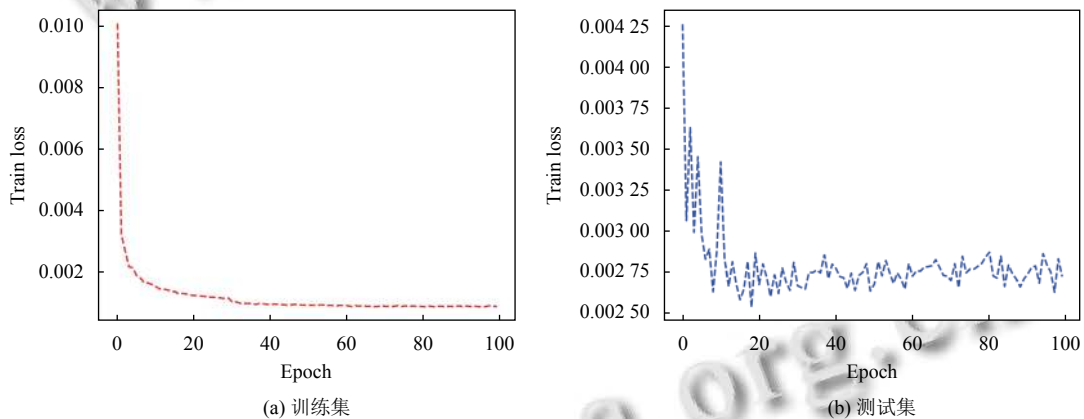


图 10 经典 U-Net 模型在训练集和测试集上的训练结果

表 2 两种网络的分割结果对比

算法	U-Net	本文算法
参数量	423 371	300 836
训练耗时 (h)	72	25
PA (%)	47.04	78.55

从表 2 可以得出本文的网络相比于传统的 U-Net 网络在性能取得提升的前提下, 大大降低了网络的参数量, 其参数量仅约为 U-Net 的 70%, 这大大提高了本文网络的训练速度, 使得本文网络的训练时间仅约为 U-Net 网络训练时间的 30%. 同时在 PA 评价指标比较下, 本文 Res-UNet 性能提高了约 60%. 综上对比结果可知, 本文所采用的网络模型相较于传统的分割网络取得显著的性能提升.

表 1 实验环境及训练主要参数

配置类型	参数值
GPU	NVIDIA RTX 2080 Ti
深度学习框架	PyTorch 1.6.0
损失函数	均方损失函数
优化器	AdaGrad

3.6 分割结果展示

在训练测试集时, 本文用预测结果对原图像进行处理, 中间黑色线段处为钢坯预测位置, 结果如图 11 所示, 精确定位了钢坯在导轨上的位置. 再将模型部署到生产环境中, 模型准确率满足钢坯在轨道上移动的定位需求.

4 结论

针对工厂生产环境中对钢坯的传输定位, 本文提出了基于 ResNet 残差块和 U-Net 的改进 Res-UNet 网络模型, 该模型在轨道钢坯定位场景中进行训练与应用的准确率达 99% 以上. 本文网络相较于传统分割网

络在性能取得提升的前提下大大减少了网络模型的训练参数数量并加快了网络的训练时间,有效降低了训练模

型的难度.实验结果表明,本文网络模型识别准确率高、速度快,能够满足复杂的生产环境中钢坯定位的需要.

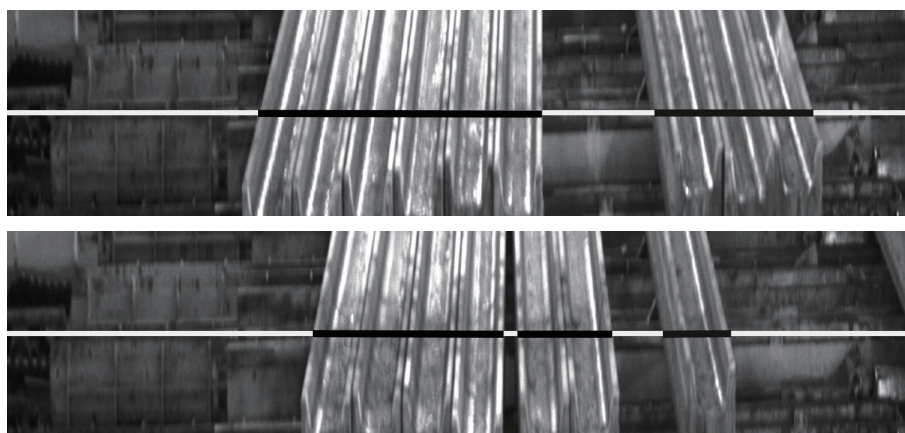


图 11 分割结果

参考文献

- 1 罗建豪, 吴建鑫. 基于深度卷积特征的细粒度图像分类研究综述. 自动化学报, 2017, 43(8): 1306–1318.
- 2 Bouthemy P, Francois E. Motion segmentation and qualitative dynamic scene analysis from an image sequence. International Journal of Computer Vision, 1993, 10(2): 157–182. [doi: 10.1007/BF01420735]
- 3 Lowe DG. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91–110. [doi: 10.1023/B:VISI.0000029664.99615.94]
- 4 张晓, 丁云峰. 基于改进 Faster-RCNN 的 IT 设备图像定位与识别. 计算机系统应用, 2021, 30(9): 288–294. [doi: 10.15888/j.cnki.csa.008077]
- 5 Everingham M, Eslami SMA, van Gool L, et al. The PASCAL visual object classes challenge: A retrospective. International Journal of Computer Vision, 2015, 111(1): 98–136. [doi: 10.1007/s11263-014-0733-5]
- 6 Uijlings JRR, van de Sande KEA, Gevers T, et al. Selective search for object recognition. International Journal of Computer Vision, 2013, 104(2): 154–171. [doi: 10.1007/s11263-013-0620-5]
- 7 Gould S, He XM. Scene understanding by labeling pixels. Communications of the ACM, 2014, 57(11): 68–77. [doi: 10.1145/2629637]
- 8 Otsu N. A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 1979, 9(1): 62–66. [doi: 10.1109/TSMC.1979.4310076]
- 9 王骏, 王士同, 邓赵红, 等. 面向小目标图像的快速核密度估计图像阈值分割算法. 自动化学报, 2012, 38(10): 1679–1689.
- 10 Lakshmi S, Sankaranarayanan DV. A study of edge detection techniques for segmentation computing approaches. International Journal of Computer Applications, 2010, CASCT(1): 35–41. [doi: 10.5120/993-25]
- 11 Adams R, Bischof L. Seeded region growing. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(6): 641–647. [doi: 10.1109/34.295913]
- 12 Hinton GE, Osindero S, Teh YW. A fast learning algorithm for deep belief nets. Neural Computation, 2006, 18(7): 1527–1554. [doi: 10.1162/neco.2006.18.7.1527]
- 13 Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe: Curran Associates Inc., 2012. 1097–1105.
- 14 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. Proceedings of the 3rd International Conference on Learning Representations. San Diego: ICLR, 2014. 1409–1556.
- 15 He KM, Zhang XY, Ren SQ, et al. Deep residual learning for image recognition. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778.
- 16 Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. Proceedings of 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich: Springer, 2015. 234–241.
- 17 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 3431–3440.

(校对责编: 牛欣悦)