文章编号：2096-1472(2021)-10-45-04
DOI:10.19644/j.cnki.issn2096-1472.2021.010.011

基于深度学习的热轧钢坯表面不同字体的字符识别研究

刘 康[1]，钱 炜[1]，杨 康[2]

(1.上海理工大学机械工程学院，上海 200093；
2.上海宝信软件股份有限公司，上海 201999)
✉1010898612@qq.com; 1458515538@qq.com; yangkang@baosight.com

摘 要：同一热轧钢坯生产线上会存在钢坯表面字符的字体不一致的问题，而利用深度学习YOLOv3算法训练不同字体的字符数据集，严重影响了整体字符的识别率，虽然原始的YOLOv3网络结构适用性较好，但对喷印字符识别区域没有针对性。为解决以上问题，根据喷印字符相对较小且没有大小形态变化的特性，改进了YOLOv3模型结构，仅保留预测小、中目标的网络结构，在保证较高检测精度的同时，缩小模型容量；采用对不同字体字符分开训练的识别方式，得出针对性分开训练比混合字体整体训练的识别准确率高的结论。结果表明，本方法比不同字体整体训练的识别准确率提高了7%以上，可在工程上进行应用。

关键词：深度学习；字符识别；热轧钢坯；YOLOv3
中图分类号：TP301.6 **文献标识码**：A

Research on Character Recognition of Different Fonts on the Surface of Hot Rolled Steel Billet based on Deep Learning

LIU Kang[1], QIAN Wei[1], YANG Kang[2]

(1.*School of Mechanical Engineering*, *University of Shanghai for Science and Technology*, *Shanghai* 200093, *China*;
2.*Shanghai Baosight Software Co.*,*Ltd.*, *Shanghai* 201999, *China*)
✉1010898612@qq.com; 1458515538@qq.com; yangkang@baosight.com

Abstract: Aiming at character fonts inconsistency on the billet surface in the same hot-rolled billet production line, deep learning YOLOv3 algorithm is used to train character data sets of different fonts, which seriously affects the overall character recognition rate. Although the original YOLOv3 network structure is quite applicable, it is not targeted at the recognition area of printed characters. In order to solve the above problem, this paper proposes to improve YOLOv3 model structure according to the characteristics of relatively small print characters and no changes in size and shape. Only the network structure for predicting small and medium targets is retained, and the model capacity was reduced while ensuring high detection accuracy. It is concluded that the recognition accuracy of the targeted separate training is higher than that of the whole training of mixed fonts. The results show that the recognition accuracy of this method is more than 7% higher than that of the whole training of different fonts, and it can be applied in engineering.

Keywords: deep learning; character recognition; hot-rolled steel billet; YOLOv3

www.rjgcz.com

1 引言(Introduction)

计算机视觉技术的迅速发展，使其得以在工业自动化生产过程中发挥着极大的推动作用，大大提高了生产效率和产品质量[1]。在钢材工件等金属工业产品生产中，每个生产工件上会采用不同的字符组成来标注其专属的生产标号，从而便于对其生产的监控、配套的管理和质量的追踪。目前，采用传统OCR技术识别字符的准确率还不理想，仍需人工读取工件上的生产标号再次确认并记录的解决方案耗费人工和时间。为实现热轧钢坯生产线达到较高的自动化水平，通过物料跟踪系统对送板、轧辊、装钢、出钢等工序进行全线数据

跟踪，其中数据跟踪发挥着至关重要的作用，而字符识别的准确率直接影响到数据跟踪[2]。车间物料跟踪的范围是从钢坯入炉，经加热炉、轧线轧制、上冷床至打捆称重为止。物料跟踪信息包括批号、轧件号、炉号、钢种、产品规格、过程数据、生产时间等，每个轧件的跟踪信息自动传输到后部工序。系统将整条轧线划分成若干个跟踪区域，对物料按照先进先出的原则实现连续实时跟踪，使实际物料在每个区域都能及时准确地显示出来[3]。因此，在复杂的生产现场环境下提高钢卷号识别率十分迫切。在现阶段研究与应用中，对于单一字体的喷印字符识别准确率比较理想，可对于混合字体的识别准确率还达不到技术要求。

如今YOLO系列的算法已有五个版本，依次为YOLOv1、YOLOv2、YOLOv3、YOLOv4、YOLOv5，但本文用到YOLOv3算法，因为YOLO系列算法中是从YOLOv3开始被广泛应用的，后面出的YOLOv4、YOLOv5是在YOLOv3基础上的进一步改进，是在细枝末节上进行的优化，没有YOLOv3在工业界的普遍适用性好[4]。该网络采用Darknet53作为骨干网络[5]，由C语言实现，容易安装，没有任何依赖项，移植性非常好；支持CPU与GPU两种计算方式，也适用于生产线现场无显卡的硬件设备；具有轻量型、灵活性的特性，适合用于来研究底层，可以更为方便地从底层对其进行改进与扩展。

因此，本文依据不同字符字体图像特征和识别任务的需求，采用基于YOLOv3的改进神经网络分开训练数据集进行对应字符识别和检测分析，旨在保证识别速度的同时，准确地实现钢坯表面喷印字符识别以便进行物料跟踪。

2 YOLOv3网络模型(YOLOv3 network model)

2.1 网络结构

我们选用Darknet53作为目标检测网络YOLOv3的骨干网络，因为其可保留原图的大部分信息，能提取待训练图像的目标特征，其中：(1)CBL由卷积层(Conv)、批量归一化(Bn)与Leaky_relu激活函数三者组成，是该网络结构中的最小组件。(2)Res unit：借鉴Resnet网络中的残差结构，将残差模块集成到网络中，得到更深的网络构建，有利于检测小目标。(3)ResX：由一个CBL和X个残差组件构成，表示这个残差块(res_block)里含有多少个Res unit，是YOLOv3网络结构中的大组件。YOLOv3的骨干网络Darknet53具有残差结构，该设计有效缓解了深层网络的梯度弥散后梯度爆炸问题，提升了检测性能，每个残差模块由2 个卷积层和1 个shortcut连接。(4)concat：张量拼接，将Darknet中间层和后面的某一层的上采样进行拼接，对应分支的特征图融合。拼接的操作和残差层add的操作是不一样的，拼接会扩充张量的维度，而add只是直接相加不会导致张量维度的改变[6]，如图1所示。

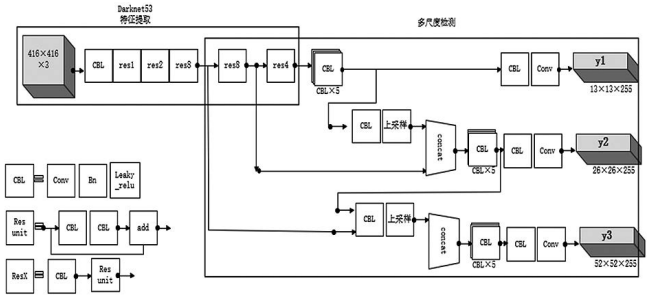

图1 YOLOv3原始网络结构

Fig.1 The original network structure of YOLOv3

2.2 网络结构的改进

原始YOLOv3网络结构采用三个不同尺度的特征图进行目标检测任务的原因是考虑原始图像中可能包含的目标物体大小不确定，对于这一问题，YOLOv3算法的解决方法是在YOLOv2曾采用passthrough层结构[7]来检测细粒度特征的基础上，对原始图像使用不同的粒度进行划分网格，更进一步采用了三个不同尺度的特征图来进行对象检测。例如，划分为13×13、26×26、52×52的网格，其中13×13的大网格用于检测相对大的目标物体，26×26的网格用于检测中等的目标物体，52×52的网格用于检测相对小的目标物体，这是与主干网络输出的三种特征图的大小是一一对应的[4]。

结合实际应用情况，钢坯表面喷印字符相对较小，字符检测任务针对的是小目标训练。YOLOv3采用多尺度来对不同尺寸的目标进行检测计算复杂度较高，对该喷印字符识别应用上有些浪费计算机算力，因此针对大小几乎无变化的喷印字符不需要过多的尺度进行训练，我们在保证较高识别准确率的同时，改进YOLOv3网络结构，仅采用两个不同尺度的特征图来进行对象检测，如图2所示。

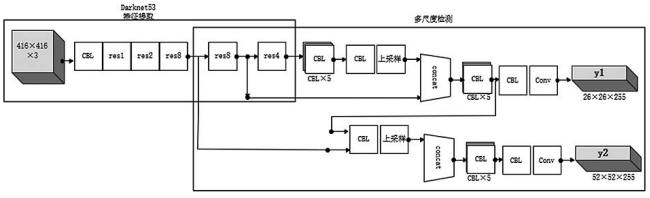

图2 改进的网络结构

Fig.2 Improved network structure

2.3 回归函数

我们知道，要实现多类别的分类，有两种改进普通的logistic回归的方式：

(1)直接根据每个类别，分别建立一个二分类器，带有这个类别的样本标记为1，带有其他类别的样本标记为0。如果有k 个类别，那么就可以得到k 个针对不同标记的普通的

www.rjgcz.com

logistic分类器。

(2)修改logistic回归的损失函数，使其适应多分类问题。这个损失函数不再只考虑二分类的损失，而是具体考虑每个样本标记的损失，这种方法就叫作Softmax回归，即logistic回归的多分类版本，可将多分类的结果以概率的形式呈现[8]。

原YOLOv3网络会产生三种不同的特征图，将被分别传入logistic层中，进而运算产生模型的输出。而logistic回归是针对二分类问题的，钢坯表面喷印字符检测识别是互斥的多分类问题，在该目标检测任务中，为了能去除无效预测框，保留最准确的预测框，同时根据喷印字符实际情况，即不存在多个目标物体重合的情况，因此普通logistic回归就不适用了，我们采用的是Softmax回归，介绍如下。

对于$\{(x_1,y_1),(x_2,y_2),\cdots,(x_m,y_m)\}$输入数据有$k$ 个类别的分类问题，先定义逻辑回归假设函数[9]，可以理解为Softmax回归估算每一类的概率，详见式(1)。

$$h_\theta\left(x^{(i)}\right)=\begin{bmatrix}p\left(y^{(i)}=1\middle|x^{(i)},\theta\right)\\p\left(y^{(i)}=2\middle|x^{(i)},\theta\right)\\\vdots\\p\left(y^{(i)}=k\middle|x^{(i)},\theta\right)\end{bmatrix}=\frac{1}{\sum_{c=1}^{k}\mathrm{e}^{\theta_c^T x^{(i)}}}\begin{bmatrix}\mathrm{e}^{\theta_1^T x^{(i)}}\\\mathrm{e}^{\theta_2^T x^{(i)}}\\\vdots\\\mathrm{e}^{\theta_k^T x^{(i)}}\end{bmatrix}\tag{1}$$

其中，$\theta_1,\theta_2,\cdots,\theta_k\in R^{n+1}$是模型的参数，$p$是类别概率值。

Softmax回归算法的代价函数[10](其中$sign(expression\ is\ true=1)$)，详见式(2)。

$$\begin{aligned}J(\theta)&=-\frac{1}{m}\sum_{i-1}^{m}\sum_{c=1}^{k}\mathrm{sign}\left(y^{(i)}=c\right)\log\left(y^{(i)}=c\middle|x^{(i)},\theta\right)\\&=-\frac{1}{m}\sum_{i-1}^{m}\sum_{c=1}^{k}1\left\{y^{(i)}=c\right\}\log\frac{\mathrm{e}^{\theta_c^T x^{(i)}}}{\sum_{l=1}^{k}\mathrm{e}^{\theta_c^T x^{(i)}}}\end{aligned}\tag{2}$$

其中，θ是回归的模型参数矩阵，c是类别，m是已标记的样本数，$1\{\cdot\}$是一个指示性函数，值为真即等于1，值为假即等于0。

通过式(2)，可将logistic回归的损失函数改为如式(3)所示。但对于$p\left(y^{(i)}=c\middle|x^{(i)},\theta\right)$，Softmax回归与logistic回归的计算方式不同，Softmax回归是logistic回归的一般形式[9]。

$$\begin{aligned}J(\theta)&=-\frac{1}{m}\sum_{c=1}^{k}y^{(i)}\log h_\theta\left(x^{(i)}\right)-\left(1-y^{(i)}\left(1-h_\theta x^{(i)}\right)\right)\\&=-\frac{1}{m}\sum_{i=1}^{m}\sum_{c=0}^{k}1\left\{y^{(i)}=c\right\}\log p\left(y^{(i)}=c\middle|x^{(i)},\theta\right)\end{aligned}\tag{3}$$

其中，θ是回归的模型参数矩阵，c是类别，m是已标记的样本数，p是类别概率值；$1\{\cdot\}$是一个指示性函数，值为真即等于1，值为假即等于0。

3　实验及结果分析(Experiment and result analysis)

3.1　实验环境

本文实验平台环境配置情况如表1所示。

表1 实验平台软硬件配置

Tab.1 Hardware and software configuration of the experimental platform

名称	相关配置
操作系统	Windows 10
中央处理器	Inter(R) Xeon(R) Golod5222 CPU
GPU	NVIDIA GeForce GTX 2080 Ti
GPU加速器	CUDA10.2　CUDNNv7.6.5
深度学习框架	Darknet53

3.2　数据集构建

图像数据集使用热轧现场拍摄作为训练识别喷印字符的数据样本，根据字体不同分开统计出两个数据集，每个数据集都包含10 种数字符号，采集图像存储格式为bmp，分辨率为2592×2048。采集的两种字体的字符图像分开整理成各自的训练数据集，两种原始字符图像如图3所示。数据集的标注采用labelimg软件，具体对图片目标区域内的单个字符依次进行标注。labelimg对字符图像标注完成后，会生成与之对应的XML文件，随后将XML文件里面的标注框名称和目标边框位置信息转换为txt文件。

(a)3 行字符图像　　(b)2 行字符图像

图3 喷印字符图像

Fig.3 Images of printed characters

3.3　训练结果

先利用本文网络模型对一种数据集进行训练，训练过程中的损失变化情况如图4所示。由图4显示的训练过程迭代次数的平均损失曲线发现，训练迭代1.5万次后平均损失函数值降低至0.35；随着迭代次数的增加，平均损失函数值基本保持不变，趋于稳定。

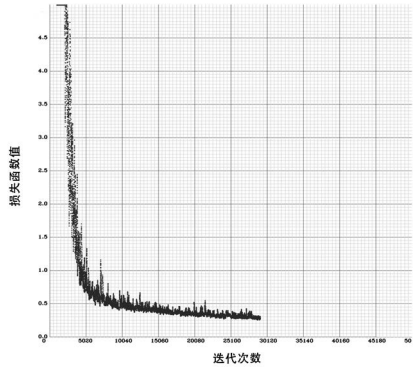

图4 训练损失变化曲线

Fig.4 Training loss curve

www.rigcz.com

分别用本文网络、原始YOLOv3网络与YOLOv3-tiny网络对相同数据集进行训练，训练好的模型测试相同测试集，结果如表2所示。

表2 三种算法测试结果

Tab.2 Test results of three algorithms

模型	原始YOLOv3网络	YOLOv3-tiny网络	本文网络
mAP	0.674	0.651	0.711
识别准确率	0.942	0.825	0.963

实验结果表明，采用本文改进的YOLOv3网络表现相对较好，在原始YOLOv3网络借鉴残差网络结构基础上，形成了更深的网络层次，去除多余多尺度检测，提升了mAP(各类别AP的平均值)及小目标检测效果。在速度相当的情况下，本文网络的识别准确率比原始YOLOv3网络与YOLOv3-tiny网络要高，得到了提升。网络的Bn层与批量大小密切相关，批量越小训练时受到的干扰越多，不同的训练批量数下损失函数收敛速度略有不同，且识别准确率也有影响，批量统计估算不准确时，在识别任务中采用小的批量数时，误差会迅速增加。用在训练大型网络和将特征转移到计算机视觉任务中时受内存消耗限制，只能使用小的训练批量数。改进的网络结构可以减小此方面的影响，提升训练批量数，减小误差。

使用本文改进的网络结构将统计的两种数据集放置一起混合训练，对混合训练与分开训练得到的模型测试相同测试集，统计结果如表3所示。显然，分开训练效果最好。

表3 不同数据集的测试结果

Tab.3 Test results of different data sets

训练集	2 行字符数据集	3 行字符数据集	混合数据集
识别准确率	0.971	0.963	0.893

4 结论(Conclusion)

本研究在Darknet53作为骨干网络的基础上，确保YOLOv3对不同尺度预测模块正常工作，结合钢坯表面喷印字符大小特征去除多尺度预测对大目标的作用，只保证对无大小变化的小目标检测无影响，减小因网络结构大而占用的计算内存，来提高批量训练的数量，在训练过程中具有很好的收敛性，训练速度也有一些提升，提升了识别准确率。网络中采用10 类别的Softmax回归层，适用多分类问题，有针对性地降低类别损失函数，提升本文网络的训练速度和识别准确率。最后证实不同字体的字符分开训练比混合训练的识别准确率更高，实际热轧现场应用分开训练模型。但本文改进的网络结构仍存在一些问题：对喷印模糊字符的图像进行特征提取效果不佳，还需提高模型的泛化能力及小样本特征提取能力。下一步准备引入图像马赛克等数据增强方法，增加样本的复杂程度，提升网络对复杂钢板表面字符图像的识别精度。

参考文献(References)

[1] 武宗茜,丁绍荣,温志强,等.巨能特钢棒材智能化生产管理系统[J].冶金自动化,2021,45(02):30-36.

[2] 王排书.热轧钢坯编号识别与表面质量检测系统研究与设计[D].锦州:辽宁工业大学,2020.

[3] 贺笛.深度学习在钢板表面缺陷与字符识别中的应用[D].北京:北京科技大学,2021.

[4] 蔡哲栋,应娜,郭春生,等.YOLOv3剪枝模型的多人姿态估计[J].中国图象图形学报,2021,26(04):837-846.

[5] 何帅.卷积神经网络在手写数字识别中的应用[J].电脑知识与技术,2020,16(21):13-15.

[6] 王辉,张帆,刘晓凤,等.基于DarkNet-53和YOLOv3的水果图像识别[J].东北师大学报(自然科学版),2020,52(4):60-65.

[7] SHEN Z J, ZANG S F, WU Q T. Weighted balanced distribution adaptation based on a softmax regression model for transfer learning[J]. Scientific Journal of Intelligent Systems Research, 2021, 3(5):27-36.

[8] 王玉,王梦佳,张伟红.基于CNN和Group Normalization的校园垃圾图像分类[J].吉林大学学报(信息科学版),2020,38(06):744-750.

[9] 徐德荣,陈秀宏,田进.稀疏自编码和Softmax回归的快速高效特征学习[J].传感器与微系统,2017,36(05):55-58.

[10] 孟佩,曹菡,师军.基于Softmax回归模型的协同过滤算法研究与应用[J].计算机技术与发展,2016,26(12):153-155,159.

作者简介：

刘 康(1996-)，男，硕士生.研究领域：机器视觉.

钱 炜(1964-)，男，硕士，副教授.研究领域：机器人机构学，机械设计及理论，CAD技术.

杨 康(1985-)，男，硕士，工程师.研究领域：智能装备，金属表面缺陷检测、分类.

www.rjgczz.com