

# 基于半监督堆叠自编码器的轧制力预报 建模研究

任巍<sup>1</sup>, 翟博豪<sup>2</sup>, 彭炜淞<sup>2</sup>

(1. 山西太钢不锈钢股份有限公司冷轧厂, 山西 太原 030003;

2. 天津电气科学研究院有限公司, 天津 300180)

**摘要:** 随着计算机运算能力的提升, 数据驱动技术被广泛应用于冶金工业过程中。基于该技术的轧制力预报有助于缩短带材的头尾长度, 提高成材率。为了解决数据驱动模型在预训练过程中因特征提取盲目导致预测精度较低的问题, 提出了一种基于半监督堆叠自编码器(SS-SAE)的深度分层监督预处理框架, 用于轧制力预报建模研究。在SS-SAE中, 依次训练多个半监督自编码器(SS-AE), 分级提取目标相关特征。每个SS-AE将来自前一隐藏层的特征作为新的输入, 以生成高阶特征。通过堆叠多个SS-AE的方式, 可逐步学习深层目标相关特征, 同时深度网络结构将逐步减少不相关信息。仿真结果表明, 该模型预测精度可控制在2%以内, 实现了轧制力的高精度预测。

**关键词:** 深度学习; 特征提取; 自编码器; 轧制力预报

**中图分类号:** TG335.13 **文献标识码:** A **DOI:** 10.19457/j.1001-2095.dqed24100

## Rolling Force Prediction Modeling Based on Semi-supervised Stacked Autoencoder

REN Wei<sup>1</sup>, ZHAI Bohao<sup>2</sup>, PENG Weisong<sup>2</sup>

(1. Cold Rolling Plant of Shanxi Taigang Stainless Steel Co., Ltd., Taiyuan 030003, Shanxi, China;

2. Tianjin Research Institute of Electric Science Co., Ltd., Tianjin 300180, China)

**Abstract:** With the improvement of computer operation level, data-driven technology is widely used in metallurgical industry. Accurate prediction of rolling force is helpful to shorten the length of strip head and tail and improve the yield of strip. In order to solve the problem of low prediction accuracy of data-driven model due to blind feature extraction during pre-training, a deep hierarchical supervised preprocessing framework based on semi-supervised stacked autoencoder(SS-SAE) was proposed for modeling rolling force prediction. In SS-SAE, several semi-supervised autoencoders(SS-AEs) were trained in turn to extract target related features. Each SS-AE took features from the previous hidden layer as new input to generate higher-order features. By stacking multiple SS-AEs, the features related to deep targets could be gradually learned, while the deep network structure will gradually reduce the irrelevant information. The simulation results show that the prediction accuracy of this model can be controlled within 2%, and the high precision prediction of rolling force is realized.

**Key words:** deep learning; feature extraction; autoencoder(AE); rolling force prediction

在现代工业过程中, 某些关键性指标的预测对安全、环保生产而言显得越来越重要, 其很大程度上决定着生产过程的规划及产品的质量<sup>[1]</sup>。近年来, 基于数据驱动的预测模型在冶金、生化工业等领域非常流行且有效, 这主要得益于大量的过程历史数据<sup>[2]</sup>。数据驱动模型无需对生产的

物理化学过程进行深入了解, 仅基于生产变量间的关系进行建模。与机理模型相比, 其具有成本低、响应快、维护方便等优点<sup>[3]</sup>。

过程数据的特征表示在建立精确的预测模型中起着不可或缺的作用<sup>[4]</sup>。由于过程数据的高维、大容量增长, 在捕获复杂数据特征方面受到

**基金项目:** 天津电气科学研究院自立项目(GY2020ZL001)

**作者简介:** 任巍(1974—), 男, 本科, 工程师, Email: renwei1@tisco.com.cn

限制。这使得传统的特征学习方法,如典型相关分析(canonical correlation analysis, CCA),主成分分析(principal component analysis, PCA)和偏最小二乘(partial least squares, PLS)等浅层学习网络模型,难以在现代大规模流程中描述数据。与浅层模型相比,多层神经网络具有更强的非线性建模能力。由于多层非线性的堆叠结构,深度神经网络能够处理任何高度非线性的数据关系<sup>[5]</sup>。深度网络主要以分层无监督预训练和有监督微调的方式进行训练,为此开发了深度信念网络(deep belief network, DBN)、长短期记忆(long and short term memory, LSTM)、卷积神经网络(convolutional neural network, CNN)、堆叠自编码器(stack autoencoder, SAE)等深度学习模型。

在以上模型中,虽然分层预训练方式可以从工业过程数据中学习到层次特征,但是大多数深度网络主要基于无监督训练的方式对原始输入数据进行特征表示,因此提取的特征可能不适合用于目标预测过程。以无监督的方式进行预训练往往会忽略引导特征学习重要参数。在实际应用中,只有从原始输入数据中提取有用的信息,才能进行高精度预测。

为此,本文提出一种基于半监督堆叠自编码器(semi-supervised stacked autoencoder, SS-SAE)的深度分层预处理框架,用于轧制力预测建模研究。将半监督自编码器(semi-supervised autoencoder, SS-AE)的编码器部分级联起来,构建一个深度网络,实现与预测目标相关的分层特征表示。通过改进特征学习的约束条件,使其在重构输入的同时,预测目标数据。然后叠加多个SS-AE模型,对原始输入数据进行连续提取,进而从低阶特征中逐层提取高阶目标相关特征。最后,将输出层添加至SS-SAE的顶层特征层,对轧制力进行预测。

## 1 半监督自编码器

### 1.1 自编码器

自编码器是一个输入层神经元数量等于输出层数量的三层神经网络。如图1所示,其网络结构分为编码层和解码层两部分。

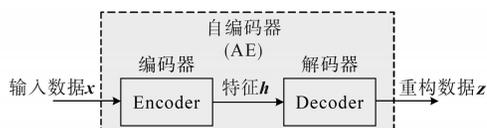


图1 自编码器模型结构

Fig.1 Structure of autoencoder

首先,输入数据 $x$ 经编码层映射至隐藏层,以提取有效特征 $h$ 。在解码层中,有效特征 $h$ 经非线性重构转换为输出数据 $z$ 。通过设定损失函数,使输出最大限度等于输入。

编码层训练过程如下:

$$h = f(x) = s(Wx + b) \quad (1)$$

式中: $W$ 为输入层到输出层的权重矩阵; $b$ 为偏置向量; $s(\cdot)$ 为非线性激活函数, sigmoid 函数。

解码层训练过程如下:

$$z = g(h) = s(W'h + b') \quad (2)$$

式中: $W'$ 为隐藏层到输出层的权重矩阵; $b'$ 为偏置向量。

损失函数为

$$l(x, z) = \frac{\sum_{n=1}^N \|x - z\|^2}{N} \quad (3)$$

式中: $N$ 为训练所需的总样本数。

### 1.2 半监督自编码器

AE是一个无监督的自学习编解码器网络。编码器将输入层映射到隐藏层的特征,解码器确保学习到的特征具有某些特定的属性。由于AE的目标是通过重构输入来学习特征,因此潜在特征需要可以在很大程度上重构其输入数据。然而,对于回归预测建模,重要的是学习有用的与预测目标相关特征。若学习具有较低相关性的特征,则会导致预测性能下降。AE作为一种无监督模型,不能保证特征与标签值的相关性,在提取特征过程中,可能存在许多不利于目标预测的无关信息。针对这一问题,设计了一种半监督自编码器(SS-AE)网络模型,用于预训练阶段学习与目标相关的特征,其网络结构如图2所示。在SS-AE网络中,使用编码器从输入数据生成隐藏特性。然后,利用隐藏特征重构输入数据的同

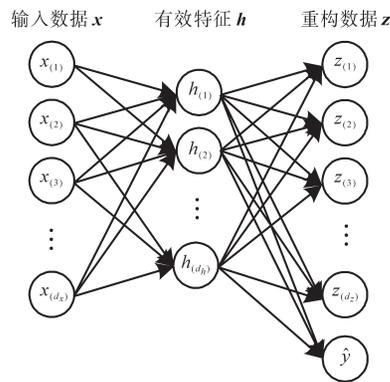


图2 半监督自编码器网络结构

Fig.2 Network structure of semi-supervised autoencoder

时,预测输出层的目标值。通过将标签值信息加入到网络中,并使用标签数据进行训练,可以提取与目标相关的特征,减少不相关信息。

由SS-AE的网络结构图可知,该模型的损失函数为

$$l(x, z) = \frac{\sum_{n=1}^N [\lambda \|x - z\|^2 + (1 - \lambda) \|y - \hat{y}\|^2]}{N} \quad (4)$$

式中: $\hat{y}$ 为目标值 $y$ 的预测值; $\lambda$ 为权重系数。

改进后的损失函数将标签值信息加入至训练过程中,使得训练过程转为半监督模式,提高了特征提取的有效性。通过调整 $\lambda$ ,可以改变标签值信息对重构过程的影响。

传统梯度下降方法中,每次迭代中使用所有样本进行训练。当训练样本较多时,训练过程较慢。小批量梯度下降将输入样本分为若干个相同大小的子集。每次迭代中,仅使用单个子集进行训练,提高了训练过程的灵活性,特别适用于大样本的训练过程。为了加快网络训练过程,SS-AE模型使用小批量梯度下降的方式进行训练,具体训练过程如图3所示。

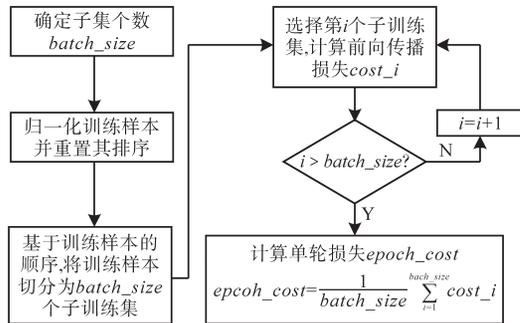


图3 单轮损失小批量梯度下降训练过程  
Fig.3 Single epoch loss function calculation of mini-batch gradient descent

Adam优化算法为模型中的每个参数提供自适应的学习率,可进一步提高训练效率,其过程如表1所示。为此,引入Adam优化算法优化模型的梯度下降过程。

### 1.3 半监督堆叠自编码器

为了逐步减少不相关信息,学习与目标相关的深层特征,将多个半监督自编码器堆叠设计,构成SS-SAE的深层网络结构。其结构如图4所示,前一个SS-AE的隐藏层作为下一个SS-AE的输入层。通过改进损失函数,将标签值信息加入到逐层的特征提取过程中。然后,学习到的深层特征可以用于目标值预测。

SS-SAE使用贪婪逐层训练的方式进行学习。

表1 Adam优化算法伪代码

Tab.1 Pseudocode of Adam optimization algorithm

Algorithm 1: Adam optimization algorithm	
1	<b>Require:</b> 学习率 $\alpha$ ,一阶和二阶矩估计的指数衰减率
2	$\beta_1, \beta_2$ , 极小数 $\epsilon$
3	<b>Require:</b> 初始化参数向量 $\theta_0$ ,一阶矩向量 $m_0 = 0$ ,
4	二阶矩向量 $v_0 = 0$ ,迭代次数 $t = 0$
5	<b>while</b> $\theta_0$ 未收敛 <b>do</b>
6	$t = t + 1$
7	计算损失函数的导数 $g_t = \nabla_{\theta} f_t(\theta_{t-1})$
8	更新一阶矩估计 $m_t = \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$
9	更新二阶矩估计 $v_t = \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$
10	计算一阶矩估计的偏差修正 $\hat{m}_t = m_t / (1 - \beta_1^t)$
11	计算二阶矩估计的偏差修正 $\hat{v}_t = v_t / (1 - \beta_2^t)$
12	更新参数向量 $\theta_t = \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$
13	<b>end while</b>
14	<b>return</b> 参数向量 $\theta_t$

对于一个 $K_m$ 层的堆叠结构SS-SAE,首先,对SS-AE<sub>1</sub>进行训练,将其输入向量 $x_1$ 映射至隐藏层 $h_1$ 中,通过最小化其损失函数 $l_1$ 提取 $x_1$ 的特征 $h_1$ 。然后,将 $h_1$ 作为SS-AE<sub>2</sub>的输入数据,用于提取高阶数据 $h_2$ 。按以上步骤,逐个训练每个SS-AE。因此,SS-SAE使用标签值变量指导每一层的特征学习,可以提取对目标值预测更有用的层次特征。

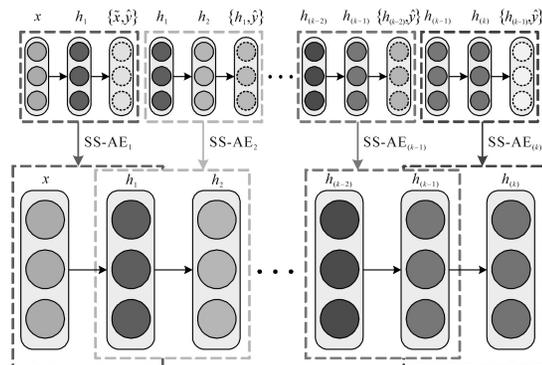


图4 半监督堆叠自编码器结构图  
Fig.4 Structure diagram of SS-SAE

## 2 基于SS-SAE的轧制力预报模型

### 2.1 基于SS-SAE的轧制力预报模型

轧制力的预报精度决定着最终成品的厚度及板形的平整度。准确的轧制力预报有助于提高成材率。基于SS-SAE的轧制力预报模型可以从原始输入数据中学习更深层次的与目标相关的特征。这些深度特征包含了大量标签值变量的预测信息,非常适合预测回归建模中。因此,只需在SS-SAE网络的最顶层增加一个用于预测

目标值的输出层,就可利用提取的深度相关特征直接构建预测模型。基于SS-SAE的轧制力预报模型详细过程描述如下:

1)根据轧制过程的机理模型确定轧制力预测建模所需的工艺变量,并划分训练集和测试集。

2)确定SS-SAE的堆叠层数,假设在SS-SAE网络中共有 $K_m$ 个隐藏特征提取层。从SS-AE<sub>1</sub>开始,使用改进的反向传播算法对其进行训练。然后,将SS-AE<sub>1</sub>的编码层变量 $\{W_1, b_1\}$ 及隐层特征向量 $h_1$ 保存下来,用于后续处理。

3)使用SS-AE<sub>1</sub>的隐层特征向量 $h_1$ 作为SS-AE<sub>2</sub>的输入数据,对SS-AE<sub>2</sub>进行训练。训练完成后,保存其编码层变量 $\{W_2, b_2\}$ 及隐层特征向量 $h_2$ 。

4)按以上方式,逐层训练每个SS-AE直至顶层。对于所有的SS-AE,其编码层变量可以被定义为 $\{W_k, b_k\}$ ,  $k = 1, 2, \dots, K_m$ 。

5)当SS-SAE特征学习网络预训练完成后,在最顶层的SS-AE隐藏层后添加目标输出层。使用标签数据,对网络中的变量 $\{W, b\}$ 进行微调。

6)在测试阶段,首先,将测试数据集 $X_i$ 输入至SS-SAE特征网络中,经第1层向顶层隐藏层的向前传播,获得目标相关的深层特征;然后,将顶层特征作为预测网络的输入,获得测试数据集的预测值 $Y_i$ 。

## 2.2 输入参数的选择

Bland-Ford 轧制力计算公式<sup>[6]</sup>为冷轧轧制力计算基本公式,具体如下所示:

$$P = Bl'_c Q_p K_1 K \quad (5)$$

式中: $P$ 为轧制力; $B$ 为原件宽度; $l'_c$ 为变形区接触弧长; $Q_p$ 为外摩擦应力状态系数; $K_1$ 为张力影响系数; $K$ 为变形抗力。

通过对Bland-Ford公式进行分析,最终选取原料宽度、原料厚度、进出口厚度、进出口张力、轧辊压扁半径和出料带线速度作为模型的输入参数。

## 2.3 结果与讨论

本模型所需样本数据选自山西太钢不锈钢股份有限公司冷轧厂的生产数据。取其轧制过程中1 000条数据作为训练集。同时,额外选择200条未参与训练的数据为测试集,以验证模型的精确性。

基于SS-SAE的轧制力预报模型分为半监督预训练和有监督微调两部分。经重复实验,最终

选定模型中的超参数如表2所示。

表2 SS-SAE模型超参数设定值  
Tab.2 Hyper-parameter of SS-SAE model

超参数	预训练阶段	微调阶段
学习率	0.02	0.01
迭代次数	200	500
权重系数 $\lambda$	0.8	—

为了获得更好的函数逼近能力,使用5个隐层数量为7的SS-AE堆叠组成轧制力预报模型,预测结果如图5所示。从图中可以看出,该模型的预测值与实测值的拟合点基本位于 $\pm 2\%$ 误差线范围内,实现了冷连轧轧制力的高精度预测,满足生产要求。

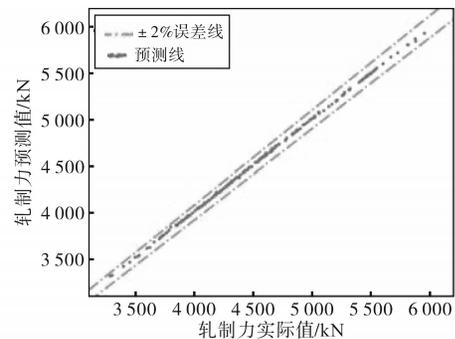


图5 基于SS-SAE的轧制力预报模型拟合效果图  
Fig.5 Simulation diagram of rolling force prediction model based on SS-SAE

为验证模型性能,将所提出的SS-SAE建模方法与其他两种相关方法进行了比较。第一种方法是传统的多层神经网络,采用反向传播算法对网络参数进行随机初始化和优化。第二种方法是基于SAE的深度网络。SAE是分层无监督模型,通过构建多个AE进行预训练。两种方法具有相同的结构,即[8,7,7,7,7,7,1]。

三种建模方法经过训练后,可以用来预测测试数据集中的目标值。表3给出了基于多层神经网络、SAE和SS-SAE预测模型的预测指标。对比均方根误差(RMSE)、平均绝对百分比误差(MAPE)和决定系数( $R^2$ ),传统多层神经网络的预测性能最差。多层神经网络虽然能够模拟复杂的非线性关系,但其参数是随机初始化的,因此,很容易进入局部最优。对于SAE,采用分层预训练技术从原始输入数据中获取深度特征,然后利用预先训练好的网络参数作为初始化参数,并对网络进行微调,因此,该方法可以提高预测精度。然而,SAE是一个无监督的自学习网络,其可以自行学习原始输入数据的良好特性不能保证特征与目标相关,在深层特征中可能包含许

多不相关的信息。与其相比,SS-SAE利用标签值构建深度半监督学习模型指导每个SS-AE的预训练。深度特征的学习是为了在每个SS-AE中尽可能好地预测目标值,这样可以把只有与质量数据相关的特征被保留下来。因此,SS-SAE可以学习深度目标相关特征,用于目标值的预测。此外,预训练的参数可以用于基于SS-SAE的深层网络的良好初始化。因此,在三种方法中,SS-SAE的预测性能最好,其RMSE,MAPE和R<sup>2</sup>指标均优于其他方法。

表3 不同模型预测效果对比

Tab.3 Comparison of prediction effects of different models

建模方法	RMSE	MAPE	R <sup>2</sup>
多层神经网络	52.457	0.835	0.893
SAE	24.699	0.513	0.975
SS-SAE	16.305	0.326	0.982

图6为三种方法在500个训练周期下的误差曲线对比。可以看出,由于多层神经网络的网络参数是随机初始化的,所以其在训练初始阶段的训练误差最大;对于SAE,其初始训练误差较小,但是SAE的预训练只是通过对输入数据的自重构来学习特征,不能保证特征与输出的相关性,因此训练误差在起始阶段较大;基于SS-SAE的预测模型在预训练阶段提取与质量相关的特征进行回归,因此SS-SAE可以获得更好的网络初始参数,并且比多层神经网络和SAE具有更小的初始训练误差,此外,与多层神经网络和SAE模型相比,SS-SAE模型可以更快地逼近收敛状态且收敛误差较低。

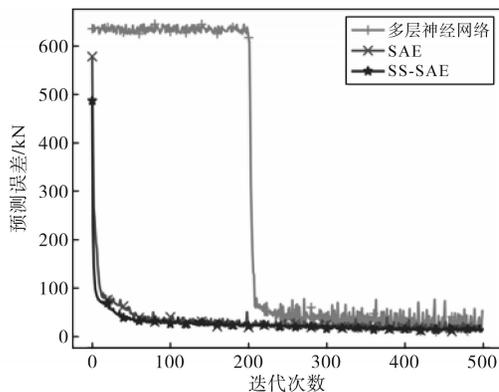


图6 不同模型训练过程

Fig.6 Training process of different models

图7比较了三种方法对测试数据集的预测误差。多层神经网络的预测误差较大。通过使用深度学习技术,SAE可以减少许多测试样本的偏差。在此基础上,SS-SAE可以学习深度质量相关

特征来预测轧制力,因此其预测误差最小。对比三种方法在测试数据集中的详细预测误差曲线可以看出,大部分的预测误差都在零附近。而且,在许多样本点上,基于SS-SAE方法的误差值均小于其他两种方法。

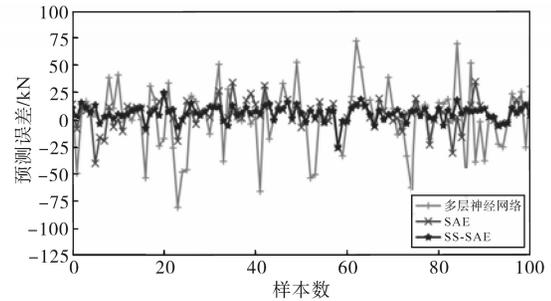


图7 不同模型对测试集的预测误差

Fig.7 Prediction error of different models for the test sample

### 3 结论

提出了一种新的半监督预测模型,并用于轧制力预测建模研究。SS-SAE模型能够从原始输入数据中学习深层与目标相关的特征。在SS-SAE中,通过构建多个SS-AE模型逐层提取深层高阶特征。与目标值相关的特征被依次学习,不相关的信息则通过层次化叠加的SS-AE逐步减少。最后,通过冷连轧生产过程的轧制力预报实验,验证了SS-SAE模型的有效性和优越性。

#### 参考文献

- 王国栋. 近年我国轧制技术的发展、现状和前景[J]. 轧钢, 2017,34(1):1-8.  
Wang Guodong. Development, present situation and prospect of rolling technology in China in recent years[J]. Steel Rolling, 2017,34(1):1-8.
- 张军,张中丹,王洲,等. 基于数据驱动的微电网双层鲁棒优化调度[J]. 电气传动,2022,52(1):68-75.  
Zhang Jun, Zhang Zhongdan, Wang Zhou, et al. Data driven two-layer robust optimal scheduling for microgrid[J]. Electric Drive, 2022,52(1):68-75.
- 任有志,乔松,孙继春,等. BP神经网络PID控制器在热油锅炉温控中的应用[J]. 电气传动,2020,50(4):81-84.  
Ren Youzhi, Qiao Song, Sun Jichun, et al. Application of BP neural network PID controller in temperature control of hot oil boiler[J]. Electric Drive, 2020,50(4):81-84.
- 杨景明,郭秋辰,孙浩,等. 基于改进果蝇算法与最小二乘支持向量机的轧制力预测算法研究[J]. 计量学报,2016,37(5):505-508.  
Yang Jingming, Guo Qiuchen, Sun Hao, et al. Research on rolling force prediction algorithm based on improved fruit fly algo-

(下转第80页)

激光雷达扫描得到的数据点集合,采用基于平行坐标系数据变换方式对数据点进行分析处理以实现集装箱的位置检测。结合起重机位置检测和定位算法,让集装箱对箱工作自动化。通过实际项目验证表明:该系统方案可行,实现了集装箱对箱的自动作业,降低了操作人员劳动强度,作业效率提高了15%。

参考文献

[1] 曾鸣. 铁路集装箱中心站物流系统资源调配优化与仿真研究[D]. 成都:西南交通大学,2015.  
Zeng Ming. Research on resources allocation optimization and simulation of railway container terminal logistics system[D]. Chengdu:Southwest Jiaotong University,2015.

[2] 梁晓波,程文明,郭鹏. 集装箱起重机自动装卸系统的研究与设计[J]. 计算机应用,2015,35(S1):229-231.  
Liang Xiaobo, Cheng Wenming, Guo Peng. Research and design of automatic handling system for container cranes[J]. Journal of Computer Applications,2015,35(S1):229-231.

[3] 杨永涛. 三维离散点云数据处理关键技术研究[D]. 秦皇岛:燕山大学,2017.  
Yang Yongtao. Research on the key technologies of three-dimensional scattered point cloud processing[D]. Qinhuangdao:Yanshan University,2017.

[4] 翟旭君. 基于平行坐标的可视化数据挖掘技术研究[D]. 北京:清华大学,2004.

Zhai Xujun. The research on visual data mining technology based on parallel coordinates[D]. Beijing:Tsinghua University,2004.

[5] 王旭宸,卢欣辰,张恒胜,等. 一种基于平行坐标系的车道线检测算法[J]. 电子科技大学学报,2018,47(3):362-367.  
Wang Xuchen, Lu Xincheng, Zhang Hengsheng, et al. A lane detection method based on parallel coordinate system[J]. Journal of University of Electronic Science and Technology of China,2018,47(3):362-367.

[6] 郑行家,钟宝江. 图像直线段检测算法综述与测评[J]. 计算机工程与应用,2019,55(17):9-19.  
Zheng Hangjia, Zhong Baojiang. Overview and evaluation of image straight line segment detection algorithms[J]. Computer Engineering and Applications,2019,55(17):9-19.

[7] 刁燕,吴晨柯,罗华,等. 基于改进的概率Hough变换的直线检测优化算法[J]. 光学学报,2018,38(8):170-178.  
Diao Yan, Wu Chenke, Luo Hua, et al. Line detection optimization algorithm based on improved probabilistic hough transform [J]. Acta Optica Sinica,2018,38(8):170-178.

[8] 赵典,强宝民,刘陈. 基于模糊自适应PID控制器的起重机精确定位技术研究[J]. 机床与液压,2013,41(7):55-58.  
Zhao Dian, Qiang Baomin, Liu Chen. Research on accurate orientation technology for gantry crane based on adaptive PID controller[J]. Machine Tool & Hydraulics,2013,41(7):55-58.

收稿日期:2020-12-03  
修改稿日期:2021-01-24

(上接第74页)

rithm and least square support vector machine[J]. Acta Metrologica Sinica,2016,37(5):505-508.

[5] 宋勇,李博,刘超,等. 基于改进堆叠自编码器的带钢力学性能预报模型[J]. 冶金自动化,2020,44(6):2-10.  
Song Yong, Li Bo, Liu Chao, et al. Prediction model of mechanical properties of strip based on improved stack autoencoder[J]. Metallurgical Automation,2020,44(6):2-10.

[6] 魏立新,翟博豪,赵志伟,等. 基于半监督深度网络的冷连轧

轧制力预报[J]. 塑性工程学报,2020,27(11):70-76.  
Wei Lixin, Zhai Bohao, Zhao Zhiwei, et al. Prediction of rolling force in tandem cold rolling based on semi-supervised deep network[J]. Journal of Plasticity Engineering,2020,27(11):70-76.

收稿日期:2021-11-22  
修改稿日期:2021-12-23