

文章编号:1671-251X(2016)12-0072-06

DOI:10.13272/j.issn.1671-251x.2016.12.016

赵燕飞,杨彦利,王丽娟.基于显著性和深度卷积神经网络的输送带表面故障定位[J].工矿自动化,2016,42(12):72-77.

基于显著性和深度卷积神经网络的 输送带表面故障定位

赵燕飞, 杨彦利, 王丽娟

(天津工业大学 天津市光电检测技术与系统重点实验室, 天津 300387)

摘要:提出了一种基于显著性和深度卷积神经网络的输送带表面故障定位方法。该方法在输送带上、下表面的边缘烙上数字,利用图像处理技术检测输送带图像中的数字来间接定位输送带表面故障。首先,将采集的输送带图像进行高斯滤波、灰度线性变换等预处理,以提高图像质量、增强背景与目标的对比度;然后根据谱残差理论对预处理后的图像进行视觉显著性处理,获取包含数字区域的视觉显著图;最后,运用卷积神经网络对显著图进行分类,以区分出数字区域和非数字区域。实验结果表明,该方法可以很好地检测输送带图像上的数字,进而实现对输送带表面故障的定位。

关键词:输送带;故障定位;神经网络;视觉显著性;图像处理

中图分类号:TD528.1

文献标志码:A

网络出版时间:2016-12-01 10:51

网络出版地址:<http://www.cnki.net/kcms/detail/32.1627.TP.20161201.1051.016.html>

Surface fault location of conveyor belt based on saliency and deep convolution neural network

ZHAO Yanfei, YANG Yanli, WANG Lijuan

(Tianjin Key Laboratory of Optoelectronic Detection Technology and Systems,
Tianjin Polytechnic University, Tianjin 300387, China)

Abstract: A surface fault location of conveyor belt based on saliency and deep convolution neural network was proposed. The method imprints figures on the edge of upper and lower surfaces of conveyor belt, and uses image processing technology to detect the number in belt image, so as to indirectly locate surface fault of the conveyor belt. Firstly, the acquired image of the conveyor belt is preprocessed by Gaussian filtering and gray-scale linear transformation to improve image quality and enhance contrast between the background and the target. Then, visual saliency treatment is conducted to the preprocessed image according to spectral residual theory, and a visual saliency map containing numeric regions is obtained. Finally, saliency map is classified by using the convolution neural network to distinguish digital region from non-digital region. The experimental results show that the method can detect number of conveyor belt image and realize surface fault location of conveyor belt.

Key words: conveyor belt; fault location; neural network; visual saliency; image processing

收稿日期:2016-07-18;修回日期:2016-10-26;责任编辑:胡娟。

基金项目:国家自然科学基金资助项目(51274150,51504164);天津市科技支撑计划重点项目(13ZCZDZX01000)。

作者简介:赵燕飞(1989—),女,山西晋城人,硕士研究生,研究方向为设备状态监测与信号处理,E-mail:zhaoyanfei226@126.com。通信作者:杨彦利(1979—),男,河北行唐人,副教授,硕士研究生导师,博士,主要研究方向为设备状态监测与信号处理,E-mail:yyl070805@163.com。

0 引言

输送带是带式输送机的重要工作部件,在众多行业应用广泛。输送带工作过程中容易发生划伤、纵向撕裂、破损等表面故障,输送带的安全可靠运行日益受到重视。随着现代科技的迅猛发展,输送带运行状态监测技术水平得到了快速提升。目前,基于机器视觉的输送带故障检测技术开始受到关注,已经取得一定成果^[1-4]。由于输送带首尾相连,循环运行,即使利用机器视觉等技术发现了其划伤等故障,在检修过程中查找故障位置仍需要耗费大量的时间、人力和物力。因此,输送带故障区域定位成为输送带视觉监测的一个重要技术问题。

针对输送带的故障定位问题,崔岩等^[5]将实时性好的CAN总线用于带式输送机运行故障定位系统,系统向各分站发送地址并接收分站发回的检测信息。洪留荣等^[6]以接头为基准参考点,提出了一种基于图像处理技术的强力输送带接头识别算法,并将其应用于强力输送带X光无损检测系统中。这些方法虽然不是针对输送带视觉检测的表面故障定位问题,但具有一定的借鉴意义。本文提出一种基于机器视觉的输送带表面故障定位方法,利用数字容易被维修人员辨认的特点,以数字为基准点,在输送带上、下表面的边缘烙上数字,采用图像处理技术检测输送带图像上的数字,从而间接定位输送带表面故障区域。

1 输送带图像数字检测流程

在实验室搭建的输送带故障视觉检测平台如图1所示。利用线阵CCD相机实时采集输送带图像,采集到的图像分为带有数字的图像和无数数字图像。若在带数字的图像上检测出故障,就可以通过该数字来对故障进行定位。由于数字以一定的间隔烙在输送带上,即使检测到没有数字的输送带图像,也可以根据采集图像的时序性,利用插值方法估计出与当前图像对应的输送带位置,从而减少因排查故障而造成的停机时间。

通过观察输送带的原始图像,发现输送带图像中数字比背景稍暗,且占整个图像的很小一部分。基于此特点,本文设计了基于视觉显著性和卷积神经网络(Convolutional Neural Networks, CNNs)的输送带图像数字检测方法,方法流程如图2所示。该方法包括3个步骤:图像预处理、粗定位和精定位。

图像预处理是为了提高图像质量,增强背景与

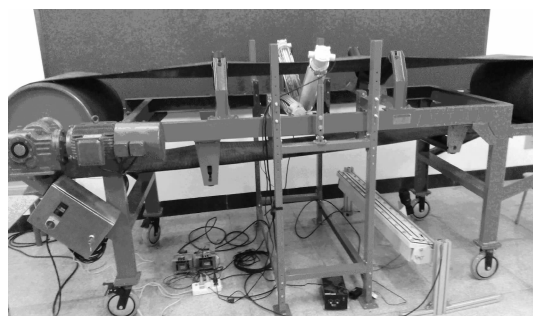


图1 输送带故障视觉检测平台

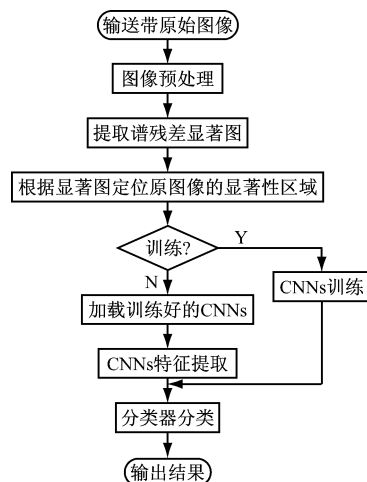


图2 输送带图像数字检测的流程

目标的对比度,以便于后续的数字识别。本文采用高斯滤波来消除噪声,平滑图像;对滤波后的图像进行灰度线性变换,以增加目标和背景的对比度。线性灰度变换函数 $L(\cdot)$ 的定义如下^[7]:

$$D_B = L(D_A) = f_A D_A + f_B \quad (1)$$

式中: D_B 表示输出图像的灰度; D_A 表示输入图像的灰度; f_A 表示线性函数的斜率; f_B 表示线性函数在 y 轴的截距。

当 $f_A > 1$ 时,输出图像的对比度将增大;当 $f_A < 1$ 时,输出图像的对比度将减小。

粗定位即利用视觉显著性对数字区域进行初步定位。利用谱残差理论对预处理后的图像进行谱残差显著性提取,获取预处理后图像的显著图,进而得到图像的视觉显著性区域。在粗定位过程中获取的显著性区域不仅包含数字,还会有一些特殊的背景。这就需要对粗定位后的图像进行进一步处理,精定位就是为此而设计的。本文运用卷积神经网络来进一步区分粗定位获取的显著性区域,首先设计卷积神经网络结构,用粗定位得到的输送带显著性区域训练卷积神经网络,得到卷积神经网络的参数,并测试网络性能,最终实现输送带数字区域的精确高效识别。

2 粗定位

观察经过预处理的输送带图像,发现数字在整个输送带图像中占很小的区域,可以将具有选择性的视觉注意机制引入输送带数字检测中,提高图像分析的效率,降低运行时间。Itti 和 Koch^[8-9]阐述了视觉显著性功能和原理。他们认为视觉显著性突出了场景中与众不同的位置,用于吸引视觉注意力。其核心是一个自下而上、刺激驱动的信号,通常是通过待识别目标与背景的不同来生成显著图。Barlow^[10]的有效编码假说和信息论相关理论表明,图像中包含的信息由特殊部分和冗余部分组成,人类的视觉系统会过滤掉冗余的部分,而剩下的特殊部分就是引起视觉注意的显著部分。Hou 等^[11]提出了谱残差方法,从分析图像频谱的角度出发,计算图像中的显著性区域。参考文献[12]指出,人类有一种对人脸与文字的定位能力,此能力在某种程度上与图像傅里叶幅度谱中的信息相关,而 Hou 等提出的谱残差显著性算法恰是基于图像傅里叶幅度谱的。因此,本文采用谱残差显著性算法对输送带数字进行定位。

首先将输送带图像 $I(x)$ 从空间域转换到频率域,得到输送带图像的幅度谱,然后对输送带图像的幅度谱取对数,得到 log 谱。获得 log 谱的过程如下^[11]:

$$A(f) = |F(I(x))| \quad (2)$$

$$P(f) = \varphi(F(I(x))) \quad (3)$$

$$L(f) = \log(A(f)) \quad (4)$$

式中: $A(f)$ 表示傅里叶变换后的幅值; $F(\cdot)$ 表示二维离散傅里叶变换; $P(f)$ 表示傅里叶变换后的相位; $\varphi(\cdot)$ 表示相位函数; $L(f)$ 表示 log 谱。

大量实验表明,不同图像 log 谱的曲线分布趋势极其相似,而且曲线满足局部线性条件^[11]。因此,用局部平均滤波器 $h_n(f)$ 对 log 谱进行平滑滤波,可获得 log 谱的大致形状。将经过平滑滤波的 log 谱作为图像的冗余信息,其表达式如下^[11]:

$$U(f) = L(f) * h_n(f) \quad (5)$$

式中: $U(f)$ 为均值滤波后的 log 谱; $h_n(f)$ 为滤波算子,是 $n \times n$ 的矩阵,定义为

$$h_n(f) = \frac{1}{n^2} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix} \quad (6)$$

将谱残差 $R(f)$ 定义为原 log 谱减去均值滤波

后的 log 谱,其表达式为

$$R(f) = L(f) - U(f) \quad (7)$$

对频率域的谱残差信息进行傅里叶反变换,得到空间域的显著图 $S(x)$,其表达式为

$$S(x) = |F^{-1}[\exp\{R(f) + iP(f)\}]|^2 \quad (8)$$

最后对 $S(x)$ 进行高斯模糊滤波就可以得到输送带原始图像的显著图。

利用实验室所采集的图像进行显著性分析,样图如图 3 所示。第 1 列为输送带原始图像;第 2 列是对应的谱残差显著性图像,图像中的亮度表示显著性程度,越亮代表越显著;第 3 列是对谱残差显著图做二值化后的图像。第 1 列的第 1 幅图像在图像的边缘包含 2 个数字 3,第 2 列第 1 幅图像中,谱残差视觉显著性算法成功抑制了输送带图像中的非显著区域,并突出了包含数字的显著区域。对谱残差显著图进行二值化处理,可清楚地看到凸显的数字 3 的区域。第 1 列的第 2 幅图像包含 2 个数字 6,第 2 列的第 2 幅图像中不仅包含数字的显著性区域,还同时保留了输送带中其他显著元素,通过精定位可有效抑制干扰元素。

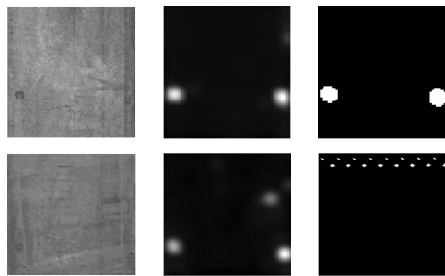


图 3 谱残差视觉显著性分析样图

3 精定位

在显著性区域中,可以分为有数字区域和非数字区域,本文采用卷积神经网络对显著性区域进行分类,去除非数字区域。

CNNs 是基于深度学习理论的一种人工神经网络,该网络可直接将图像数据作为输入,避免了图像预处理和特征提取等复杂操作^[13],被广泛应用于图像识别领域。它不仅可显著提升识别准确率,而且在特征提取速度方面也优于许多传统方法^[14-15]。CNNs 是一种多层神经网络,由卷积层和下采样层交替组成,每层又分为多个平面,而每个平面又包含多个独立神经元。关于 CNNs 的理论推导见参考文献[16]。

本文采用 6 层网络,CNNs 由 2 个卷积层、2 个下采样层交替组成,如图 4 所示。

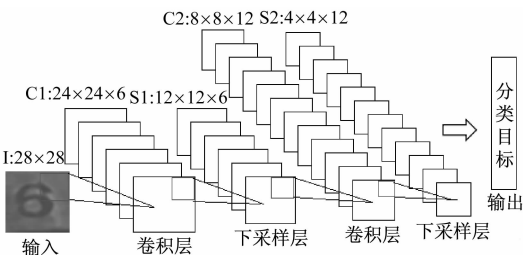


图4 CNNs 结构

有关检测输送带数字的 CNNs 框架的详细描述如下。

输入层 I: CNNs 能够对二维图像的特征进行自主学习,在确保输入图像细节信息的情况下,为减少数据量,将通过显著性得到的区域进行灰度化,通过插值计算变化为 28×28 大小的图像,将该图像归一化后作为网络的输入数据。

C1 层是特征提取层,用 6 个 5×5 大小的卷积核分别去卷积 28×28 大小的输入图像,最终获得 6 个大小为 $(28-5+1) \times (28-5+1) = 24 \times 24$ 的特征图。卷积核大小的选取对于提高 CNNs 的性能至关重要。若卷积核过小,则无法有效提取局部特征;若卷积核过大,则提取的局部特征的复杂度可能远远超过卷积核的表示能力^[17]。同一个特征图使用的卷积核是一样的。由于 Sigmoid 函数有良好的非线性映射特征,所以将 Sigmoid 激活函数作用于卷积后的结果作为 C1 层某个神经元的特征值,存储在 C1 层。C1 层需要训练的参数数目为 $6 \times (5 \times 5 + 1) = 156$ 个,其中的 1 代表偏置项。输入层和 C1 层的连接数为 $156 \times 24 \times 24 = 89\ 856$ 。

S1 层是下采样层,其目的是减小 C1 层特征图的大小。对 C1 层中非重叠的 2×2 大小的区域求和,再乘以一个权重,加上一个偏置,最终获得 6 个 12×12 大小的特征图。为防止丢失更多图像细节特征,选用 2×2 大小的下采样域。每个子采样特征图需要训练 2 个参数,S1 层共有 $6 \times 2 = 12$ 个参数需要训练。

C2 层也是一个特征提取层,最终获得 12 个 8×8 大小的特征图。所用卷积核大小跟 C1 层相同,但所获的特征图个数不同。C2 层需要将 S1 层的 6 个特征图进行组合,映射出 12 个特征图。本文让 CNNs 网络自己学习组合。

S2 层是子采样层,它获得 12 个 4×4 大小的特征图,过程类似于 S1 层。

输出层是一个全连接层。S2 层共有 $12 \times 4 \times$

$4 = 192$ 个神经元,输出层共有 2 个神经元(即输送带数字区域和非数字区域),一个神经元与输出的每个神经元相连,所以共有 $192 \times 2 = 384$ 个连接。将 S2 层看作一个 192 维的线性向量,则 S2 层到输出层的映射等同于使用该线性向量进行分类,分类器共有 384 个参数。

4 实验与结果分析

为了验证本文方法的有效性,在 Matlab 平台上进行实验,所有原始数据均采集于实验室中的输送带。

4.1 粗定位结果与分析

由于参考文献[18]与本文粗定位阶段研究的问题类似,所以采用参考文献[18]提出的评价标准:

$$RR = \frac{DTB}{TB} \tag{9}$$

$$RP = \frac{DTB}{DTB + DNB} \tag{10}$$

$$RF = 2 \times RP \times \frac{RR}{RP + RR} \tag{11}$$

式中:RR(Region Recall)为检测到的数字区域召回率;DTB 为检测到的数字区域个数;TB 为真实数字区域的个数;RP(Region Precision)代表检测到的数字区域准确率;DNB 为检测到的特殊背景区域的个数;RF 为综合 RR 和 RP 的算法整体性能。

参考文献[19]比较了 4 种显著性算法,认为 Torralba 的显著性算法最适用于文字检测。参考文献[20]将 Torralba 的显著性算法与谱残差算法比较,综合运行速度等因素,认为谱残差视觉显著性算法更适用于文字检测。本文方法是在参考文献[20]的基础上,结合输送带图像的特点,先对输送带图像进行预处理,再采用谱残差视觉显著性算法得到输送带数字的视觉显著性图。用含有 0—9 数字的 10 幅输送带图像进行实验,10 幅图像中共有 22 个数字。按照式(9)一式(11)的评价标准,将本文方法与参考文献[20]中提出的方法进行比较,结果见表 1。

表 1 显著性算法比较			%
方法	RP	RR	RF
参考文献[20]方法	12.5	36.4	18.6
本文方法	46.7	95.5	62.7

从表 1 可以看出,本文方法的准确率 RP、召回率 RR 和整体性能 RF 都要比参考文献[20]中的方

法高。召回率 RR 高表明本文方法检测的数字区域更完整,遗漏的数字区域更少;准确率 RP 高意味着检测到的数字区域比背景区域要多;整体性能 RF 是对上面 2 个指标的综合,越高越好。

从 10 幅输送带图中选出 3 幅图作为例子,对比结果如图 5 所示。其中第 1 列是原始图像,第 2 列是用本文方法得到的显著图,第 3 列是其二值图;第 4 列是用参考文献[20]中的方法得到的显著图,第 5 列是其二值图。

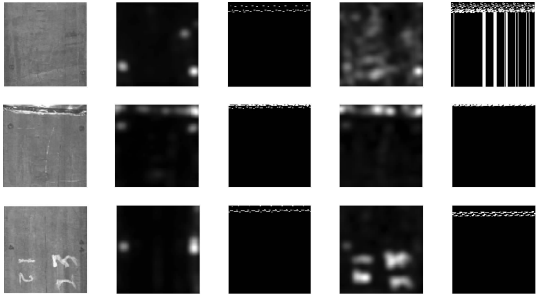


图 5 显著结果对比

从图 5 可以看出,由于本文方法考虑了输送带图像中滤波和数字与背景的对比度,检测出的数字区域更完整,会出现少量的背景区域。参考文献[20]中的方法不仅会出现数字区域不完整的情况,还会出现大量的背景区域。

综上所述,本文方法相比参考文献[20]中的方法具有更好的性能。由于是粗定位,所以需要更高的召回率 RR,以利于后续算法的整体性能提升。本文方法的召回率在只有 22 个数字区域的情况下达到 95.5%,但准确率为 46.7%,表明定位结果检测的数字区域很完整,同时检测到的非数字区域也很多。因此,需要对数字区域进一步定位。

4.2 精定位结果与分析

部分输送带图像如图 6 所示。为了确保实验的准确性和可靠性,本文选择 258 幅 $1\,024\times1\,024$ 大小的输送带图像,200 幅用于训练集,58 幅用于测试集。训练时,首先用谱残差获得视觉显著图,然后定位到原图的显著性区域进行切割,切割出的图像包括数字区域和非数字区域,选择 400 个正样本数据(数字区域)和 800 个负样本数据(非数字区域)。由于图 4 所示的 CNNs 只接收 28×28 大小的灰度图像,所以,首先需要对切割的图片进行灰度化处理;然后,对图片尺寸进行缩放。由于实验室采集的输送带图像是在同一光照下,尺度视角基本没有发生变化,所以,先使用简单算法进行归一化,利用双线

性插值进行几何缩放,将所有图片调整到 28×28 之后,再输入到 CNNs。测试时的正样本是 116 个,负样本是 234 个。



(a) 原始图像



(b) 数字区域(正样本)



(c) 背景区域(负样本)

图 6 部分输送带图像

在神经网络训练中一个重要的参数是训练的迭代次数。迭代次数过小会导致网络的预测误差过大,对输入无法充分学习,导致最终多层的深度学习算法的分类准确度过低;迭代次数过大会导致计算时间过长,得到的准确率只有微小的提升。表 2 显示了迭代次数与识别率的关系,识别率是在迭代次数一样的情况下,取 5 次训练的平均结果。从表 2 可以看到,随着迭代次数的增加,准确率会趋于稳定,并最终收敛。为了减少实验的计算时间,在实验中取迭代次数为 100,此时的准确率较高且计算时间也较短。从表 2 可看出,神经网络对输送带数字区域检测的正确率可达到 99.91%。神经网络能够以较高的可靠性区分出输送带图像的数字区域和非数字区域,平均耗时为 0.001 3 s。

表 2 迭代次数与识别率的关系

迭代次数	识别率/%	
	训练样本集	测试样本集
20	94.90	91.94
40	98.17	97.83
60	99.37	99.24
80	99.78	99.78
100	99.91	99.91

用 4.1 节中的评价标准,取 4.1 节中所用的 10 幅图像来评价显著性+卷积神经网络方法的性能,结果见表 3。从表 3 可以看出,进行视觉显著性提取后再进行卷积神经网络分类,输送带数字区域的识别准确率 RP 由最初的 46.7%增加到 100%,

充分说明后期的分类效果很好。整体性能 RF 从 62.7%提升到 97.7%。准确率 RP 与综合性能 RF 的提升说明本文方法对输送带数字的定位更准确。由于输送带的运行速度较快,留给每帧图像的处理时间较短,用本文方法处理每幅图像的平均时间 t 为 0.271 3 s,可知该方法运行速度快,可以满足实时性要求。

表 3 显著性+卷积神经网络方法的性能

RP/%	RR/%	RF/%	t/s
100	95.5	97.7	0.271 3

5 结语

针对输送带视觉检测的表面故障定位问题,利用数字容易被维修人员辨认的特点,以数字为基准点,提出了一种基于机器视觉的输送带表面故障定位方法。运用谱残差理论对输送带图像进行了处理,理论分析和实验结果表明,视觉显著性可以突出输送带图像中的数字区域,但同时也会将若干非数字区域突出显示。利用 CNNs 对经过显著性提取后的输送带图像进行分类,以区分出数字区域和非数字区域。实验结果表明,CNNs 可以很好地区分数字区域和非数字区域。利用实验室现有的输送带进行测试,发现该方法精度高,定位准确且速度快,满足实时性要求。

参考文献:

[1] 张晞,刘鸿鹏,叶涛.带式输送机纵向撕裂数字图像检测系统设计研究[J].煤炭工程,2011,43(10):16-18.

[2] 唐艳同,乔铁柱,牛犇.输送带纵向撕裂在线监测预警系统的设计[J].煤矿机械,2012,33(5):242-244.

[3] YANG Y,MIAO C,LI X,et al. On-line conveyor belt sinspection based on machine vision [J]. Optik-International Journal for Light and Electron Optics, 2014,125(19):5803-5807.

[4] 亢亢,苗长云,杨彦利.基于显著性和最大熵的输送带撕裂检测[J].计算机系统应用,2013,22(3):117-120.

[5] 崔岩.基于CAN总线长距离胶带输送机故障定位系统的研究[D].沈阳:沈阳工业大学,2010.

[6] 洪留荣.强力输送带接头识别算法[J].工矿自动化,2012,38(4):9-12.

[7] 张铮,王艳平,薛桂香.数字图像处理与机器视觉:

Visual C++与 Matlab 实现[M].北京:人民邮电出版社,2010.

[8] ITTI L, KOCH C, NIEBUR E. A model of saliency-based visual attention for rapid scene analysis [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998, 20(11): 1254-1259.

[9] ITTI L, KOCH C. Computational modelling of visual attention[J]. Nat Rev Neurosci,2001,2(3):194-203.

[10] BARLOW H B. Possible principles underlying the transformation of sensory message [M/OL]// ROSENBLITH W A. Sensory communication. MIT Press Scholarship Online; 2013 [2016-05-25]. DOI: 10.7551/mitpress/9780262518420.003.0013.

[11] HOU X D, ZHANG L Q. Saliency detection: a spectral residual approach[C]//IEEE Conference on Computer Vision and Pattern Recognition, 2007: 1-8.

[12] HONEY C, KIRCHNER H, VAN RULLEN R. Faces in the cloud: Fourier power spectrum biases ultrarapid face detection[J]. Journal of Vision, 2008, 8(12): 1-13.

[13] 赵志宏,杨绍普,马增强.基于卷积神经网络 LeNet-5 的车牌字符识别研究[J].系统仿真学报,2010,22(3):638-641.

[14] 余永维,殷国富,殷鹰,等.基于深度学习网络的射线图像缺陷识别方法[J].仪器仪表学报,2014,35(9):2012-2019.

[15] 高学,王有旺.基于CNN和随机弹性形变的相似手写汉字识别[J].华南理工大学学报(自然科学版),2014,42(1):72-76.

[16] BOUVRIE J. Notes on convolutional neural networks [EB/OL]. (2006-11-22) [2016-06-06]. http://cogprints.org/5869/1/cnn_tutorial.pdf.

[17] 龚丁禧,曹长荣.基于卷积神经网络的植物叶片分类[J].计算机与现代化,2014(4):12-15.

[18] SHI C, XIAO B, WANG C, et al. Graph-based background suppression for scene text detection[C]//10th IAPR International Workshop on Document Analysis Systems, Gold Coast,2012: 210-214.

[19] SHAHAB A, SHAFAIT F, DENGEL A. How salient is scene text [C]//10th IAPR International Workshop on Document Analysis Systems, Gold Coast,2012: 317-321.

[20] 姜维,卢朝阳,李静,等.基于视觉显著性与文字置信图的场景文字的背景抑制方法[J].电子学报,2015,43(1):62-68.