

DOI:10.3969/j.issn.1001-4551.2018.09.020

基于机器视觉和神经网络的低质量文本识别研究

李少辉,周 军*,刘 波,钱俞好,吴闽仪
(河海大学 机电工程学院,江苏 常州 213000)

摘要:针对流水线产品上的文本图片含有较多噪声和缺陷,造成机器视觉产品中字符识别准确率低,鲁棒性差的问题,对文本识别中的图像预处理、字符切分和归一化、字符识别等方面进行了研究,采用机器视觉技术中基于仿射变换的预处理方法对文本图片进行了倾斜校正,保证了后续字符的精确切分;提出了一种基于改进的 BP 神经网络算法,显著提高了字符识别的准确率和鲁棒性,利用附加动量法和自适应学习速率方法,避免了传统 BP 神经网络训练过程中易陷入局部极小值的情况,并提高了神经网络模型的收敛速度。研究表明:该方法能使倾斜的图片得到有效校正,神经网络模型的训练效率得到有效提高,且在含有噪声和缺陷的字符图片中仍能保持较高的识别率和鲁棒性。

关键词:机器视觉;流水线;文本识别;倾斜校正;改进 BP 神经网络

中图分类号:TP242.6⁺2;TH39

文献标志码:A

文章编号:1001-4551(2018)09-1006-05

Low quality text recognition based on machine vision and neural network

LI Shao-hui, ZHOU Jun, LIU Bo, QIAN Yu-hao, WU Min-yi

(School of Mechanical and Electrical Engineering, Hohai University, Changzhou 213000, China)

Abstract: Aiming at the problem of the text picture on the assembly line product contains more noises and defects, which results in the low accuracy of the character recognition and the poor robustness of the machine vision product, the image preprocessing, character segmentation and normalization, character recognition in the text recognition were studied, By using the preprocessing method based on affine transformation in machine vision technology, the text image was skewed to ensure the accurate segmentation of subsequent characters. An improved BP neural network algorithm was proposed to improve the accuracy and robustness of character recognition. By using the additional momentum method and the adaptive learning rate method, the traditional BP neural network was easy to fall into local minima and the convergence speed of neural network model was improved. The results show that the proposed method can effectively correct the skew image and maintain high recognition rate and robustness in the character picture with noise and defect.

Key words: machine vision; assembly line; text recognition; tilt correction; improved BP neural network

0 引 言

近年来,为了进一步提升生产车间流水线的自动化水平,提高生产效率,大部分流水线开始采用机器视觉产品代替人眼来识别产品上的文本信息,如生产日期、产品批号、条形码等^[1]。传统的文字识别分类器有模板匹配、支持向量机等,早在 20 世纪 80 年代,

OCR 技术就已进入商用阶段,但是对文字背景的要求很高,需要很好的成像质量;近年来,随着人工神经网络技术在很多问题上有着突破性进展,该技术也被引入文本识别领域中。2014 年 ICPR 会议上,微软亚洲研究院团队训练的一个浅层 BP 神经网络在自然场景文字检测的标准数据集上取得了 92.1% 的检测精度,让自然场景图像中的文字检测实现了突破^[2];谷歌公

收稿日期:2018-01-04

作者简介:李少辉(1994-),男,河南许昌人,硕士研究生,主要从事机器视觉和机器学习方面的研究。E-mail:hohai_lsh@163.com

通信联系人:周 军,男,博士,教授,硕士生导师。E-mail:zhj6171@163.com

司 2015 年发布的 tensorflow 深度学习框架中,深度卷积神经网络对手写数字图片集 MNIST 的识别准确率达到了 99% 以上^[3]。

由于大部分产品的文本信息是印刷体字符,包含较多的噪声和缺陷,且文本背景比较复杂,本文采用传统的图像处理方法和一种改进 BP 神经网络对流水线产品上的文本信息进行识别。

1 图片的预处理

图片预处理过程包括彩色图像灰度化、二值化、倾斜矫正等^[4]。由于产品的文本图片由于放置问题可能出现倾斜现象,对后续的字符分割和识别产生影响,本文提出了基于仿射变换的倾斜校正方法。

1.1 彩色图像灰度化和二值化

大部分彩色图片采用 RGB 系统,所包含的信息量巨大。为了减少运算量,需要将彩色图像转换为灰度图像,即图像的每一个像素颜色 RGB 各个分量的值都相等^[5]。

灰度化处理,将图像二值化变成只有黑和白两色图像的过程称为二值化处理。二值化可以在简单的场景中通过阈值化处理得到。阈值化是输入图像 f 到输出图像 g 的变换,即:

$$g(i, j) = \begin{cases} 1 & \text{当 } f(i, j) \geq T \text{ 时} \\ 0 & \text{当 } f(i, j) < T \text{ 时} \end{cases} \quad (1)$$

式中: T — 阈值; $g(i, j)$ — 二值化后的图像像素; $f(i, j)$ — 原始图像像素,当 $f(i, j) \geq T$ 时为前景图像,否则为背景图像。

在工业环境光照不均匀的情况下图像的灰度变化较大,整幅图像不存在合适的单一阈值,因此可采用多阈值法进行二值化处理,但其算法时间和空间复杂度较大。本文先采用均值滤波的方法对图像做整体滤波处理,然后使用 Ostu 自适应阈值法选取二值化阈值。其基本思想是:对于图像 $f(i, j)$,前景(即目标)和背景的分割阈值记作 T ,则类间方差可表述为:

$$g = \omega_0 \omega_1 (\mu_0 - \mu_1)^2 \quad (2)$$

式中: g — 类间方差; ω_0 — 前景的像素点数占整幅图像的比例; ω_1 — 背景像素点数占整幅图像的比例; μ_0 — 前景的像素点的平均灰度; μ_1 — 背景的像素点的平均灰度。

采用遍历的方法得到使类间方差 g 最大的阈值 T ,即为所求阈值。

流水线上某产品的标签图像,以及采用自适应阈值法将标签二值化后的图像如图 1 所示。

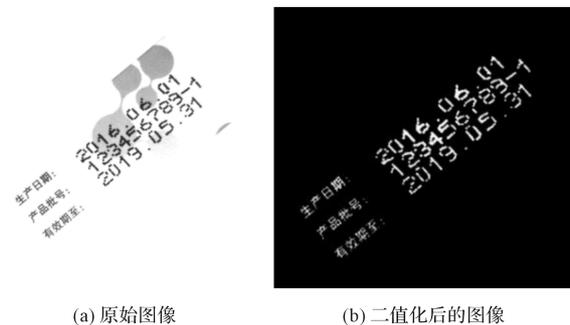


图 1 原始图像和二值化后的图像

由图 1 可以很清楚地看到白色区域是文本区域。

1.2 倾斜矫正

图片倾斜会影响后续字符的分割和识别^[6],图片倾斜校正的过程如图 2 所示。

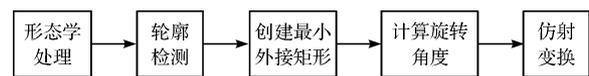


图 2 倾斜校正流程

1.2.1 形态学处理与轮廓检测

形态学变换膨胀采用向量加法对两个集合进行合并^[7]:

$$X \oplus B = \{P \in \mathcal{E}^2, P = x + b, x \in X \text{ 且 } b \in B\} \quad (3)$$

式中: X — 原始图像; B — 膨胀元素结构; P — 合并后的图像。

将二值化后的图像分别采用 $30 \times 20, 50 \times 40$ 的结构元素膨胀,结果如图 3 所示。

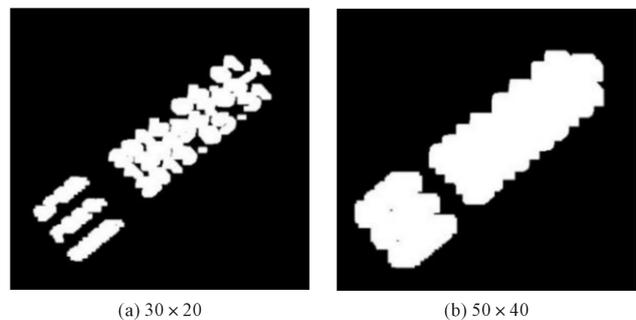


图 3 采用不同膨胀核膨胀后的结果

图像处理中用 canny, sobel 等算子进行图像的边缘检测^[7],该过程通过图片中像素值的不同可以将图片的边界检测出来,但在处理中,并没有把边界被当做一个整体。轮廓检测过程正是将所检测出来的边缘像

素进行整合,将其组装在一起。具体实现过程如下:扫描图像中的全部像素,当扫描到连通域中的一个像素点时,从该像素点开始,对边界上的像素点进行记录,当边界轮廓是一个整体的时候,回到前一个位置,进行新的扫描过程。

1.2.2 创建最小包围矩形并计算倾斜角度

为了获取文本区域的倾斜角度,需要做轮廓的最小外接矩形。最小包围矩形的宽、高及倾斜角度如图 4 所示。

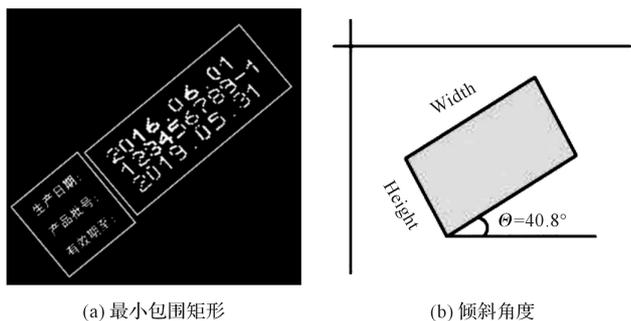


图 4 最小包围矩形及其倾斜角度

1.2.3 仿射变换

仿射变换^[8]保持了二维图形的“平直性”和“平行性”。本研究采用的变换形式是旋转。变换后的图像坐标为:

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (4)$$

式中: θ —图像倾斜角度; $\begin{bmatrix} x \\ y \end{bmatrix}$ —原始图像。

旋转后的图像如图 5 所示。



图 5 旋转后的图像

2 字符切分与归一化

文字的切分是文字识别的很重要的一个环节^[9]。目前主要的切分方法有经典切分法、连通区域切分法、基于识别的切分法、投影法等。这里采用投

影法。

图像的投影一般情况下是沿着垂直或者水平的方向,即沿着列或者行的方向,这样得到的投影被称作为竖直和水平投影。图像 $I(x,y)$ 在 x 轴和 y 轴上的投影表示为:

$$\text{垂直投影: } I(x) = \sum_{y=0}^{W-1} I(x,y), (0 \leq x \leq H) \quad (5)$$

$$\text{水平投影: } I(y) = \sum_{x=0}^{H-1} I(x,y), (0 \leq y \leq W) \quad (6)$$

式中: $I(x)$ —垂直投影图; $I(y)$ —水平投影图; $I(x,y)$ —原始图像; W —图像宽度; H —图像的高。

笔者利用水平投影将文本的每一行分割出来,然后再采用垂直投影将每个字符分割开来。为了将分割出来的字符尺寸统一到标准模板大小,需要将字符进行归一化处理^[10]。

3 字符识别

目前用于字符识别的方法已经有很多种,如模板匹配、支持向量机算法等。由于大多数流水线产品上的印书体字符噪点和缺陷较多,以上几种文字方法识别准确率较低。BP 神经网络^[11]结构简单、可操控性好,因此在实际应用中广受欢迎。

3.1 BP 神经网络模型的训练

笔者将每一个字符种类的模板库作为神经网络的一个输入神经元,建立三层 BP 神经网络模型,如图 6 所示。

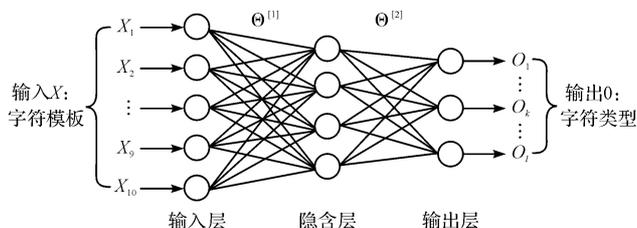


图 6 神经网络模型

输入层到隐含层采用 RELU 非线性激活函数为;

$$X_2 = \max(0, \theta_1 X_1 + b) \quad (7)$$

式中: X_1 —输入层向量 $[x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(m)}]$; θ_1 —输入层到隐含层的权值向量 $[\theta_1^{(1)}, \theta_1^{(2)}, \dots, \theta_1^{(n)}]$; b —偏置项; X_2 —隐含层向量 $[x_1^{(1)}, x_1^{(2)}, \dots, x_1^{(k)}]$ 。

隐含层到输出层采用 softmax 激活函数为:

$$h(x) = \frac{e^{-\theta_2^{(i)} T X_2}}{\sum_{i=1} e^{-\theta_2^{(i)} T X_2}} \quad (8)$$

式中: $h(x)$ —训练获得的输出层结果; θ_2 —隐含层到输出层权值。

由代价函数获得输出层误差,即:

$$\min_{\theta} J(\theta) = -\frac{1}{m} \left[\sum_{i=1}^m y_i \log(h(x_i)) + (1 - y_i) \log(1 - h(x_i)) \right] \quad (9)$$

式中: $J(\theta)$ —训练误差; m —样本数量; y —网络输出值。

反向传播采用梯度下降法不断更新权值,即:

$$\text{repeat} \left\{ \theta_j := \theta_j - \eta \frac{\partial}{\partial \theta_j} J(\theta) \right\} \quad (10)$$

式中: η —学习速率; θ_j —最小误差下的权值($j = 0, 1, \dots, n$)。

3.2 BP 神经网络的改进

3.2.1 网络模型收敛速度的提高

在 BP 神经网络参数中,学习速率 $\eta(0 < \eta < 1)$ 的取值直接影响到学习网络的收敛速度。为了提高学习速度,应选用较大的 η 值,而 η 太大却可能导致在稳定点附近振荡,甚至不收敛。因此,学习速率必须小于某一上限,这就使得 BP 算法的收敛速度受到影响。

自适应学习速率算法通过检查网络权值的修正值是否降低了误差函数来自动调节学习速率大小,使得网络总是以最大的可接受学习速率进行训练^[12]。当观测到连续两次训练的网络误差值减小时,则增大学习速率,以期缩短学习时间;此时,若误差的反弹在允许的范围,则保持学习速率;而当误差增大时,则及时减小学习速率值,直至网络误差收敛为止。常见的自适应学习速率调整公式为:

$$\eta(t+1) = \begin{cases} \beta_1 \cdot \eta(t), & E(t) < E(t-1) \\ \beta_2 \cdot \eta(t), & E(t) > \xi \cdot E(t-1) \\ \eta(t), & \text{其他} \end{cases} \quad (11)$$

式中: β_1, β_2 —学习速率调整比例因子; ξ —允许反弹误差系数; $E(t)$ —网络误差; $\beta_1 > 1, 0 < \beta_2 < 1, \xi$ 通常略大于 1。

3.2.2 附加动量法

BP 神经网络算法的反向传播实质上是采用随机梯度下降法,从随机起点沿代价函数梯度方向逐渐使代价函数的值达到最小的过程。代价函数为一个凹凸不平的多维空间的超曲面,存在较多的局部极小值点,

因此训练网络时可能陷入某一极小值,使训练出的网络模型失去泛化能力。

附加动量法是在每个权值的变化上加上一项正比于前次权值变化的值,并根据反向传播法产生新的权值变化。该方法使模型在更新梯度时,不仅考虑了误差在梯度上的作用,而且考虑了误差曲面上变化趋势的影响。带有动量因子的权值更新公式为:

$$\begin{cases} \Delta \theta_i^{(j)}(t+1) = (1 - \alpha) \eta J(\theta) P_j + \alpha \Delta \theta_i^{(j)}(t) \\ \Delta b_i(t+1) = (1 - \alpha) \eta J(\theta) + \alpha \Delta b_i(t) \end{cases} \quad (12)$$

式中: t —训练次数; $\Delta \theta_i^{(j)}$ —权值变化量; α —动量因子; P_j —输入矢量; Δb_i —偏差变化量。

3.3 实验结果分析

本研究以 Visual studio 和 Opencv 为实验平台,采用三层改进 BP 神经网络对 5 000 张字符图片(含 100 种字符,每种字符含 50 张不同的图片)进行学习,其中隐含层神经元个数为 128,学习速率取 $\eta = 0.01$,动量因子 α 取 0.9,学习速率衰减系数取 0.1,最终模型的训练精度为 94.6%,测试精度为 93.8%。本研究调用训练好的神经网络模型,将训练集中没有出现过的新字符图片输入训练好的模型中。无噪声和缺陷的字符图片在不同识别方法下的识别正确率和模型的训练时间如表 1 所示。

表 1 无噪声和缺陷的字符图片识别结果

识别方法	字符总数(无噪声和缺陷)	正确识别字符数	正确率 /%	模型训练时间/min
BP 神经网络	500	458	91.6	12.5
改进的 BP 神经网络	500	481	96.2	10.8
模板匹配	500	342	70.0	/
支持向量机	500	401	80.2	20.3

有噪声和缺陷的字符图片中不同识别方法识别正确率和模型的训练时间如表 2 所示。

表 2 有噪声和缺陷的字符图片识别结果

分类器	字符总数(有噪声和缺陷)	正确识别字符数	正确率 /%
BP 神经网络	300	255	85
改进的 BP 神经网络	300	282	94
模板匹配	300	123	41
支持向量机	300	224	74.7

对比分析实验结果:表 1 表明改进的神经网络方法对于训练集的依赖性较小,适用于训练集中没有出现的新图片。传统的模板匹配方法对模板的依赖度较高,对新图片的适应能力较弱;而基于支持向量机的方法在训练大型样本时,时间代价高且准确率较低,与之相比,本文采用的方法有较强的鲁棒性和识别率;表 2 表明传统的 BP 神经网络泛化能力较弱,在含有噪声和缺陷的字符图片中识别准确率明显降低,而改进的 BP 神经网络不易陷入局部极小值,使网络具有了一定的稳定性,在含有噪声和缺陷的低质量文本图片中仍能保持较高的识别准确率,且有效缩短了网络模型的训练时间。

4 结束语

本文详细论述了一种基于改进 BP 神经网络的流水线产品上的低质量文本识别的全过程,包括了图像的预处理、文本字符的切分与归一化和字符识别方法等,主要结果有:

(1)解决了由于图片倾斜而不能进行精确字符分割的问题;

(2)使用学习速率自适应调整策略和附加动量法改进了 BP 神经网络,显著提高了网络模型的识别准确率并有效缩短了神经网络模型的训练时间。与传统的模板匹配、支持向量机和传统的 BP 神经网络相比,模型的泛化能力和鲁棒性较强。

参考文献 (References):

- [1] 王焯青,杨永跃. 机器视觉在流水线条形码识别中的应用[J]. 电子测量与仪器学报,2006,20(6):102-105.
- [2] 杨 飞. 自然场景图像中的文字检测综述[J]. 电子设计工程,2016,24(24):165-168.
- [3] 莫宇琨,信昆仑,陈 能. 深度卷积神经网络在水表字符图像识别上的应用[J]. 供水技术,2017,11(5):54-57.
- [4] 刘明周,马 靖,张 森,等. 基于机器视觉的机械产品装配系统在线作业方法[J]. 计算机集成制造系统,2015,21(9):2343-2353.
- [5] 张 丛. 基于机器视觉的机械零件的识别研究[D]. 天津:天津职业技术师范大学机电工程学院,2015.
- [6] 毕文波. 基于机器视觉的机械零件尺寸识别研究[D]. 济南:山东大学机械工程学院,2006.
- [7] 黄 山. 基于机器视觉的机械臂分拣系统的研究[D]. 镇江:江苏科技大学电子信息学院,2016.
- [8] 秦旭光,徐爱功. 图像几何校正在 OCR 文字识别中的应用[J]. 辽宁工程技术大学学报,2006(S2):64-66.
- [9] 唐 璁,李 青. 一种快速的模板匹配算法[J]. 计算机应用,2010,30(6):1559-1561,1564.
- [10] ZHENG H, XIE L, ZHANG L Z. Sensitivity Analysis of Electricity Price Based on Support Vector Machine [J]. **Proceedings of the CSEE**,2006,26(11):134-138
- [11] GONG T, HU T S, TIAN X Z. Human face expression recognition based on within-class modular PCA [J]. **Mechanical & Electrical Engineering Magazine**,2009,26(7):74-76.
- [12] 许宜申,顾济华,陶 智,等. 基于改进 BP 神经网络的手写字符识别[J]. 通信技术,2011,44(5):106-109,118.

[编辑:李 辉]

本文引用格式:

李少辉,周 军,刘 波,等. 基于机器视觉和神经网络的低质量文本识别研究[J]. 机电工程,2018,35(9):1006-1010.

LI Shao-hui, ZHOU Jun, LIU Bo, et al. Low quality text recognition based on machine vision and neural network[J]. *Journal of Mechanical & Electrical Engineering*, 2018,35(9):1006-1010.

《机电工程》杂志;http://www.meem.com.cn