

基于可变形卷积神经网络的数字仪表识别方法



郭兰英 韩睿之 程鑫

长安大学信息工程学院 西安 710064

(lyguo@chd.edu.cn)

摘要 目前,对于数显仪表的识别,多采用传统的图像处理及机器学习等方法,在复杂多变的应用场景中,其对字符、数字的识别准确率低,难以满足实时应用的要求。针对以上问题,将传统图像处理技术与深度学习方法相结合,提出了一种基于可变形卷积神经网络的数显仪表示数分割与识别方法。该方法包含图像预处理、字符分割与识别等步骤。首先,使用 GrayWorld 算法对待识别图像进行亮度均衡,并通过彩色分割提取屏幕区域;其次,对图像进行形态学操作,以便使用投影直方图法完成字符与对应小数点的整体分割;最后,设计并训练了一种可变形卷积神经网络对字符进行识别,优化了卷积神经网络感受野几何结构固定的内在问题。实验结果表明,加入可变形卷积有效提高了图像的识别准确率和网络的收敛速度;该方法的整体识别准确率达到 99.45%,检测速度为 10FPS,能够满足实际应用需求。

关键词: 图像处理;字符识别;可变形卷积神经网络;投影直方图

中图分类号 TP391.4

Digital Instrument Identification Method Based on Deformable Convolutional Neural Network

GUO Lan-ying, HAN Rui-zhi and CHENG Xin

School of Information Engineering, Chang'an University, Xi'an 710064, China

Abstract At present, traditional image processing methods and machine learning methods are adopted for the identification of digital display instruments, which have disadvantages such as low recognition accuracy for both characters and numbers in complicated scenarios, and difficulty to meet real-time application requirements. Aiming at the problems above, combining traditional image processing technology and deep learning methods, a method of segmentation and recognition of digital display instrument based on deformable convolutional neural network is proposed. This method includes steps such as image preprocessing, character segmentation and image recognition. Firstly, the GrayWorld algorithm is applied to perform brightness equalization on the image to be recognized for the further using of color segmentation to extract the screen area. Secondly, the projected histogram method is implemented to realize the unified segmentation of characters with its corresponding decimal point after performing morphological operation on the image. Finally, a deformable convolutional neural network is proposed and trained for character recognition, which optimizes the endogenous geometry restriction of receptive field in convolutional neural networks. The experimental results indicate that the addition of deformable convolution effectively improves the accuracy of image recognition and the convergence speed of the network, and the accuracy of the overall recognition method reaches 99.45% and the detection speed is 10FPS, which can meet the requirements of practical applications.

Keywords Image processing, Character recognition, Deformable convolutional neural network, Projection histogram

1 引言

随着科学技术的发展,各行业的管理方式正在由人工管

理方式逐渐变为自动或半自动管理方式,工业的自动化也是发展的必然趋势。汽车尾气分析仪在交通监管与环境保护中有着广泛的应用,对于检测汽车尾气排放、降低废弃污染有着

收稿日期:2019-10-09 返修日期:2020-03-12 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:陕西省重点研发计划(2019NY-163);陕西省交通科技项目(14-23K);中央高校基本科研业务费专项资金项目(300102329101, 310824175004)

This work was supported by the Shaanxi Provincial Key Research and Development Program(2019NY-163), Shaanxi Provincial Transportation Science and Technology Project(14-23K) and Central University Basic Research Business Expenses Special Fund Project(300102329101, 310824175004).

通信作者:韩睿之(ruizhih@qq.com)

极为重要的作用。目前,尾气分析仪的读数仍然依靠人工实现,不仅效率较低,还有着较高的错误率。为了解决以上这些问题,可以使用计算机视觉的方法对仪器数据进行识别,同时使得监测数据的记录和传输更加方便。若要实现检测的自动化,则需要使用数码相机实时拍摄尾气分析仪器的显示区域,通过计算机程序完成图像的处理和识别,将图像信息转化为文本信息。因此,图像识别与处理算法具有巨大的实用价值,对数字仪表的自动监测具有重要的作用和意义。

一般地,针对仪表液晶屏幕的识别数字仪表的识别流程主要有3个部分,分别为图像预处理、字符分割和字符识别。其中,预处理主要包括图像二值化、去噪、倾斜矫正等操作;字符分割以图像预处理为基础,对后续的识别准确度产生了重要的影响,常用的分割方法有投影直方图分割^[1]以及基于联通区域的MSER等;字符识别是整个流程的关键部分,早期主要使用模板匹配的方法,目前以特征提取加分类器分类为主。

模板匹配是一种最基本的图像识别方法,通过将模板在目标图像上平移来比较模板与其覆盖范围的相似度,确定特定目标在图像中的位置,从而进行图像的识别。随后,针对字符识别,研究者采用尺度不变的特征变换^[2](Scale-invariant Feature Transform, SIFT)以及方向梯度直方图^[3](Histogram of Oriented Gradient, HOG)等算子进行图像的特征提取,并在特征空间上使用KNN(K-Nearest Neighbors)、SVM(Support Vector Machines)等机器学习算法构建分类器,进而进行图像的分类识别。然而,这一阶段的图像识别系统多针对某一特定任务而设计,且泛化能力受到数据量的影响,在实际应用中难以获得精准的识别效果。

在2012年的ILSVRC^[4-5]比赛中,Alex等基于LeNet^[6-7]设计了一种面向大规模数据的卷积神经网络AlexNet^[8],该网络首次使用了ReLU函数作为激活函数,对卷积输出进行非线性化,以及使用权重随机失活(DropOut)技术进行参数正则化等。其后,卷积神经网络^[9]以自主提取图像特征的能力,在图像分类^[10]、目标检测^[11-12]和实例分割^[13-14]等多种视觉识别任务中获得了广泛的应用。其中,GooLeNet^[15]通过网络结构的嵌套和堆叠提高了模型的表达能力,并且引入了Inception^[16]模块和BottleNeck层,从而大大减小了算法的运算量;VGG16^[17]通过减小卷积核尺寸但增加卷积层数的方式,在减少计算参数的同时保证了感受野的大小,并且通过更深的网络结构获得了更好的非线性表达能力。然而,层数的机械性增加会使网络出现退化问题(Degradation problem),即网络深度增加到一定程度后,准确率出现饱和甚至下降的现象。因此,He等在ResNet^[18]中使用了identity mapping和residualmapping两种方式进行前向传播,以使网络保持在最优状态,从而避免了深层网络带来的训练难度和准确率问题。

然而,上述的卷积网络都只通过最大池化^[19]等降采样模块来获得一定的形变不变性,其效果极其微弱,且受制于池化单元的大小。为了进一步提升卷积神经网络的空间不变性,可变形部件模型(Deformable Part Model, DPM)^[20]基于

HOG算子对两种精度的图像进行梯度检测,并将物体建模成多个部件的组合,设定启发式规则进行目标判定;空间变换网络(Spatial Transformer Network, STN)^[21]尝试在神经网络中增加定位网络和网格生成器,显式地对特征图进行空间映射来学习目标的几何形状。而Luo等^[22]通过研究发现,感受野中各位置对输出的影响并不相等,而是呈高斯分布。根据这一发现,Yu等^[23]提出了膨胀卷积,通过对卷积核的各个元素增加大于1的步长来有效增加感受野的尺寸,同时保留原有的计算复杂度;Dai等^[24]提出了可变形卷积(Deformable Convolutional Network, DCN),通过对每个卷积核的位置增加一个二维的偏移量,使得卷积核可以在规则网格周围随机采样,不同位置的卷积核可以根据图像内容自适应物体的形状、大小而产生变化,削弱了标准卷积由于规则采样对网络几何形变适应性的影响。

基于以上研究,本文将可变形卷积应用到数字仪表的卷积神经网络识别^[25]中,从而使算法快速适应分类目标之间的几何形变,取得了更好的分类准确率;针对汽车尾气分析仪示数排版具有一定规则的特点,提出了一种将小数点与数字进行整体识别的预处理方法。该方法通过彩色分割及仿射变换对汽车尾气分析仪显示屏区域进行提取和矫正,并使用纵向膨胀操作和投影直方图分割法对示数及检测项字符进行预处理,最终基于可变形卷积设计了一种卷积神经网络以完成字符图像的识别。

2 基于可变形卷积神经网络的数字仪表识别

2.1 可变形卷积神经网络

在视觉识别任务中,通常采用增强数据集或基于形变不变性特征的算法来适应目标的形态差异,这两种方法存在着相同的缺陷,即只能适应已知且固定类型的形变,对于复杂的形变则难以适应。卷积神经网络的数据采样主要使用规则的格点对特征图的固定位置进行采样,对于几何形变的建模能力主要来自数据集的扩增、较大的网络层数以及模型的人为设计,仍未能从根本上解决以上问题。

为了解决卷积神经网络对目标物体形状多样性的不适应问题,本文引入了可变形卷积和可变形ROI池化的思想,通过对传统卷积和池化操作的采样点添加二维偏置值,来实现卷积核、池化框采样格点的自由形变,从而使卷积神经网络适应目标物体的不规则形状。其中,卷积核的偏置值通过额外的卷积层来获得,而池化框的偏置则由特征图和ROI共同提取,以自适应地实现对不同形状物体各部分的定位。偏置获取的模型都为简单的层,因此这一过程所需的参数和计算量都非常小,并且可以通过梯度反向传播算法来实现端对端的训练。

传统卷积使用规则网格 R 在输入特征图固定位置进行采样,并使用权重 w 对采样值进行加权求和,网格 R 定义了感受野和膨胀参数。一般地,尺寸为 3×3 、膨胀参数为1的卷积核 R 的定义如下:

$$R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\} \quad (1)$$

则对于输出的特征图 y ,其某一位置 p_0 的值为:

$$y(p_0) = \sum_{p_n \in R} x(p_0 + p_n) \cdot \omega(p_n) \quad (2)$$

其中, p_n 为采样格点中的位置。在可变形卷积中,使用偏置序列 $\{\Delta p_n | n=1, 2, \dots, N\}$ 来增强卷积核的各个采样点 p_n 的位置,此时卷积的计算方式变为:

$$y(p_0) = \sum_{p_n \in R} x(p_0 + p_n + \Delta p_n) \cdot \omega(p_n) \quad (3)$$

其中, N 为采样点个数,当前卷积核在每个位置上的偏置 Δp_n 通过单独的卷积层得到。

通过单独的卷积层获取采样偏置的结构示意图如图 1 所示,其实现方式为:在输入到输出特征图的计算中,先通过上部分的卷积层计算代表特征图两个方向的二维偏置,在输入上进行偏置映射,再对添加了偏置的输入进行普通卷积得到输出特征图。其中,偏置映射结果与输入的特征图尺寸相同,获取偏置的卷积核大小与同一层的卷积核相同,两个卷积核的权重同时更新。

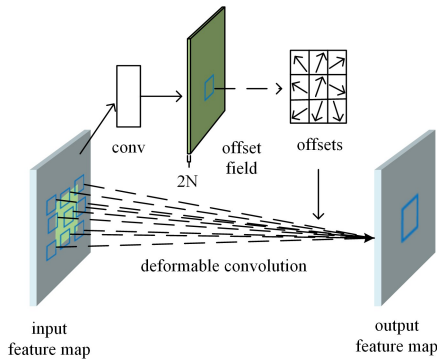


图 1 可变形卷积示意图

Fig. 1 Diagram of deformable convolution

由于通过卷积层得到的像素偏置值往往为小数,因此需要通过线性插值来获取实际的像素偏移位置。将两个一维的线性插值核 $g(a, b) = \max(0, 1 - |a - b|)$ 分别应用于特征图的垂直方向和水平方向,以实现双线性插值,则双线性插值核表示为:

$$G(p, q) = g(q_x, p_x) \cdot g(q_y, p_y) \quad (4)$$

其中, p 表示可变形卷积公式(见式(3))中的任意小数位置 $p_0 + p_n + \Delta p_n$ 。

ROI 池化用于将任意大小的感兴趣区域进行划分和降采样,输出固定大小的特征集合,在基于区域生成的目标检测方法中得到了广泛的应用。与可变形卷积类似,可变形池化对每个池化框的位置添加二维偏置值,通过添加同一层的卷积和全连接来输出每个池化框的偏置位移。对于输入特征图 x 中左上角位置为 p_0 、宽和高分别为任意 w 和 h 的感兴趣区域,可变形池化将其划分为 $k \times k$ 个池化框进行平均池化,并固定输出长度为 $k \times k$ 的特征图 y ,且 y 中第 (i, j) ($0 \leq i, j \leq k$) 个元素的计算方式为:

$$y(i, j) = \sum_{p \in \text{bin}(i, j)} x(p_0 + p + \Delta p_{ij}) / n_{ij} \quad (5)$$

其中, p 为第 (i, j) 个池化框中的像素位置, n_{ij} 为该池化框中包含的像素个数。在可变形池化中,首先通过 ROI 池化生成输出特征图,在此特征图上添加全连接层获取归一化的池化

偏置,将此偏置与输入的宽、高进行元素级相乘得到实际偏置,并使用式(4)的双线性插值将实际偏置映射到输入特征图上进行实际的池化操作。

2.2 图像数据预处理

2.2.1 屏幕区域矫正

由于显示屏幕具有反光的特性,仪表图像极易受到阳光、灯光等光照因素的影响,造成图片的过亮或过暗,因此需要对屏幕区域进行提取和矫正,对图像进行亮度均衡。本文使用 GrayWorld 算法完成亮度平衡,该算法假设自然界景物对于光线的平均反射均值在总体上为定值,该值可根据图像 3 个通道的亮度平均值求得,并用于计算每个通道的增益系数,从而对图像进行亮度均衡。GrayWorld 算法的流程如算法 1 所示。

算法 1 GrayWorld 算法

输入:原图像

输出:经过亮度均衡化的图像

1. 求反射均值,即图像 3 个色彩通道亮度的平均值:

$$G_{\text{ray}} = (\bar{R} + \bar{G} + \bar{B}) / 3 \quad (6)$$

其中, $\bar{R}, \bar{G}, \bar{B}$ 为彩色图像在 R, G, B 3 个色彩通道上的亮度平均值;

2. 根据反射均值计算 3 个通道的增益系数 k_r, k_g, k_b , 其计算方式为:

$$\begin{cases} k_r = G_{\text{ray}} / \bar{R} \\ k_g = G_{\text{ray}} / \bar{G} \\ k_b = G_{\text{ray}} / \bar{B} \end{cases} \quad (7)$$

3. 根据增益系数 k_r, k_g, k_b 调整每个像素 C 的通道的值,使得:

$$\begin{cases} C(R') = C(R) \times k_r \\ C(G') = C(G) \times k_g \\ C(B') = C(B) \times k_b \end{cases} \quad (8)$$

根据 GrayWorld 算法得到平衡之后的图像,避免因光照等因素造成成像效果的差异。将均衡后的图片转换为 HSI 色彩空间,并通过色调值 H 过滤非屏幕区域。通过多次实验发现,汽车尾气分析仪屏幕的背景色调值一般为 105~110,而饱和度与亮度值的变化范围较大。因此,本文按照表 1 中的数据设定阈值对图像进行彩色分割,得到屏幕位置指示图像,记为 I_{ind} 。

表 1 彩色分割参数

Table 1 Parameters used for color segmentation

	Min_value	Max_value
Hue	100	115
Saturation	120	200
Intensity	210	260

在得到屏幕区域位置 I_{ind} 后,便可通过仿射变换得到矫正的屏幕图像 I_{scr} 。按照面积大小对轮廓排序,通过宽高比确认屏幕区域,在图像 I_{ind} 中查找屏幕边缘轮廓的最小外接矩形 MER 并求得 4 个顶点,从而通过仿射变换得到宽、高分别为 800 像素、600 像素的屏幕图像 I_{scr} 。本文所使用的汽车尾气分析仪屏幕的宽高比例 w/h 为 4:3。

将经过 GrayWorld 算法进行亮度均衡后提取的屏幕图像 I_{scr} 与原始输入图像进行对比,效果如图 2 所示。

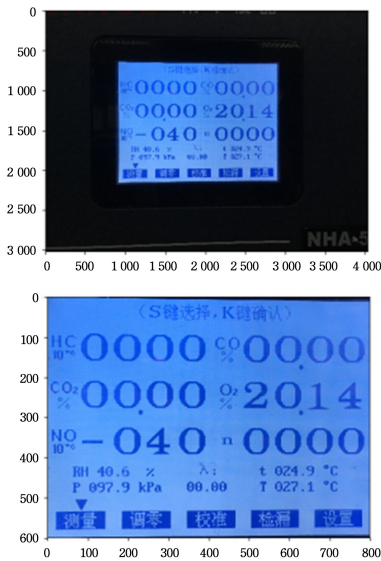


图2 屏幕区域的提取

Fig. 2 Screen extraction

2.2.2 字符预处理与分割

对提取的屏幕图像进行预处理,以得到更好的字符分割效果,预处理主要包含二值图像的滤波以及膨胀操作,字符分割使用投影分割法。使用高斯滤波对屏幕二值图像进行去噪后发现,每个小数点与其对应的数字在纵向投影上相连通,但在横向投影上存在间隔,在字符分割时造成分离,因此本文使用纵向膨胀操作对图像进行进一步处理,以提高图像的分割效果。

图像的膨胀是一种常见的形态学操作,通过将物体接触的背景点合并到该物体中,使得物体的边界向外部扩张。具体地,膨胀操作以一个结构元素 B 为模板在二值图像上进行滑窗操作,将 B 与其覆盖的二值图像范围做“或”操作,若结果为 1 则将部分图像值置为 1,否则置 0。其定义形式如下:

$$E = F \oplus S = \{(x, y) | S_{xy} \subseteq X\} \quad (9)$$

本文使用规格为 5×3 的菱形结构元素对图像进行纵向膨胀,以弥合小数点与数字之间的纵向间隔,使得小数点与对应数字被分割在同一字符图像中。膨胀效果如图 3 所示。经过预处理的屏幕图像消除了图像中存在的噪点,弥合了小数点区域与对应数字的垂直间隔,能够通过直方图投影法得到较好的分割结果。本文对经过纵向膨胀的图像进行横向扫描,获得横向的分割起始点对,对原屏幕二值图进行纵向扫描,得到纵向分割起始点对,从而将两者组合为每个字符的

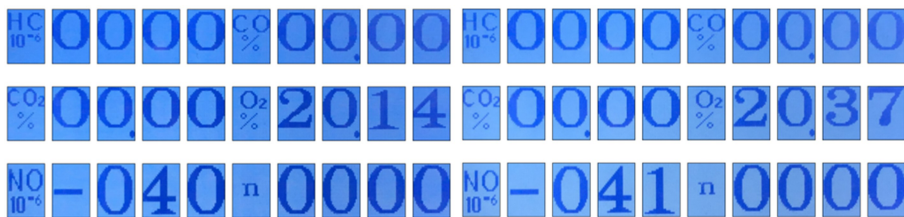


图4 投影直方图法分割的字符图像矩阵

Fig. 4 Character image matrix segmented by projection histogram

ROI(Region of Interest)位置。将小数点与其对应的数字分割于同一字符图像中,有助于下一步的整体字符识别,降低了小数点的错误分割,同时提高了识别准确率。

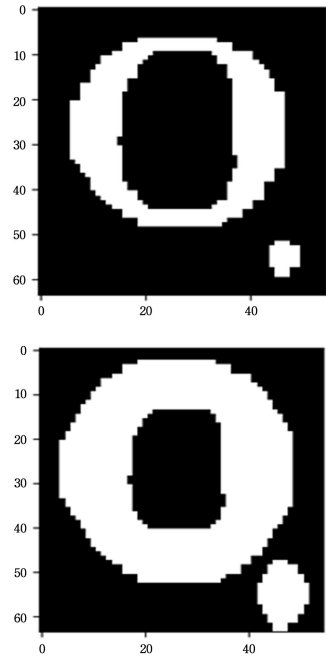


图3 纵向膨胀效果图

Fig. 3 Result of longitudinal dilation

将屏幕区域二值图像记为 I ,经过预处理的图像记为 I_b ,则通过直方图投影法进行字符分割的具体步骤如算法 2 所示。通过直方图投影法分割得到的字符图像矩阵如图 4 所示。

算法 2 直方图投影分割法

输入:二值屏幕图像 I_b

输出:字符图像矩阵

1. 对二值图像 I_b 进行横向投影,即根据计算式 $Num_i = \sum_{j=0}^w I_b[i][j]$ 统计每行的前景像素个数 Num_i ,其中 $i=1,2 \dots h$ 为图像每行的索引, w 表示图像 I_b 的宽度。
2. 计算横向分割点对,即遍历 Num_i 记录元素全为 0 的子集的起始位置,记为 $[s_1, e_1], [s_2, e_2], \dots, [s_m, e_m]$,其中 m 为分割得到的行图像个数;通过分割点对并按行对图像 I 进行分割,得到行图像序列 I_{h1}, I_{h2}, \dots 。
3. 对每个行图像 I_{hi} 进行纵向投影,得到纵向投影直方图并进行纵向分割,将分割的字符图像按行、列号保存为二维图片矩阵,记作 I_{vj} ,其中 $i=1,2, \dots, m, j=1,2, \dots, n$, m 和 n 分别为图像中字符的行数和列数。

2.3 网络结构与损失函数

2.3.1 网络结构

本文利用可变形卷积的思想,基于 LeNet5 网络的经典结构,将普通卷积层后的池化层改为使用 ReLU 作为激活函数的激活层,并使用 BatchNormalization 对每一个 mini-batch 进行归一化,同时通过调整卷积步长的方式实现降采样。在原本的两个卷积层后添加两个包含可变形特性的可变形卷积层,即先使用 ConvOffset2D 层将卷积核位置偏移映射到特征

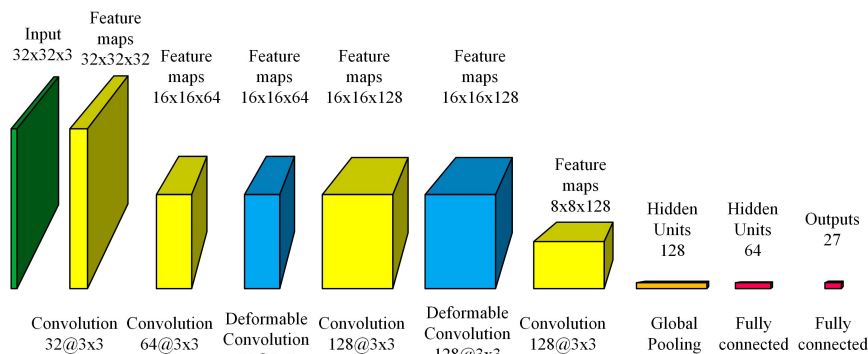


图5 可变形卷积神经网络结构图(电子版为彩色)

Fig. 5 Structure of deformable convolutional neural network

2.3.2 损失函数

交叉熵作为一种概率距离的度量方式,常在分类问题中作为损失函数对模型进行评估。对于二分类问题,使用二值交叉熵 binary_crossentropy 损失函数,其定义如下:

$$\text{loss} = -\sum_{i=1}^n y_i \log y_i + (1 - y_i) \log(1 - y_i) \quad (10)$$

其中, y 代表真实值, \hat{y} 代表预测值, n 为样本个数。当预测值与真实值相等时,损失函数值为 0,两者相差越大,损失值也越大。

本文中,对于分割的字符图片集合,需要将其正确地识别为对应的示数或检测项。作为多分类问题,在经过卷积层完成特征提取后连接 softmax 函数以输出类别分布概率,使用分类交叉熵 categorical_crossentropy 计算损失值用于比较预测值与真实值的差距,从而优化模型的参数。分类交叉熵的计算方式如下:

$$\text{loss} = -\sum_{i=1}^n y_{i1} \log y_{i1} + y_{i2} \log y_{i2} + \dots + y_{im} \log y_{im} \quad (11)$$

其中, m 为数据类别数目。

2.4 识别结果解析

将屏幕图像分割的字符图像存储于二维矩阵中,按照行优先顺序输入神经网络进行识别,将代表识别结果的数字保存于二维矩阵中,记为 R_{ij} ,其中 $i=1,2,\dots,m,j=1,2,\dots,n$,矩阵中的元素为 0 到 26 的数字。通过算法 3 解析结果矩阵,以得到正确的语义信息。

算法 3 识别结果解析算法

输入:结果矩阵 R_{ij}

输出:代表各检测项及检测值的语义信息

1. 对于矩阵 R_{ij} 的每一行 R_i ,扫描元素值以确定检测项的位置,记为 $p_0, p_1 \dots p_l$,其中 l 为该行检测项的个数。
2. 对于每个检测项 k ,其检测示数保存于 R_{i,p_k+1} 到 R_{i,p_k+1-1} 中,记为

图上,再对映射特征图进行卷积操作;经过卷积层完成特征提取后,对每个特征图进行全局池化并取得全局特征,将结果输入全连接层进行分类。

本文提出的可变形卷积神经网络的具体结构如图 5 所示,其中黄色块代表普通卷积层,蓝色块代表可变形卷积层,27 代表字符图像的种类个数。图 5 省略了卷积层之后的激活层和 BatchNormalization 归一化层,其输入图像的大小归一为 32×32 像素。

$N[0,1,2,\dots,t]$,对于实验采用的仪器,有 $t=4$,采用式(12)计算中间量 s :

$$s = \sum_{i=0}^n N_i \times 10^{n-i} \quad (12)$$

3. 在 $N[0,1,2,\dots,t]$ 中查找包含小数点的数字位置,即大于 9 的元素的位置 j ,若 $s \geq 10^{n+1-j}$,则令 $s = s - 10^{n+1-j}$;若 j 不存在,则执行下一步。
4. 对于调整后的中间值,采用式(13)计算小数点尾数,所得结果 c 即为尾气分析仪器中指标的检测读数:

$$c = s / 10^{n-j} \quad (13)$$

针对图 4 中第一个分割字符图像矩阵,输入神经网络进行识别的结果及解析的正确语义信息如图 6 所示。图 6(a)为可变形卷积神经网络输出的识别结果矩阵,图 6(b)为通过本文方法分析得到的语义信息。

23	0	0	0	0	21	0	10	0	0
22	0	10	0	0	26	2	10	1	1
25	20	0	4	0	24	0	0	0	0

(a)神经网络输出

HC	0000	CO	00.00
CO ₂	00.00	O ₂	20.11
NO	-040	n	0000

(b)系统识别结果

图 6 识别输出及检测结果分析

Fig. 6 Recognition results and analysis

3 实验及结果分析

3.1 数据与实验

实验基于 OpenCV 和 Keras 深度学习框架完成,数据来源于对南华某型号汽车尾气分析仪所采集的屏幕图像,共得到字符图像 2400 张,包含多种拍摄角度、清晰度及光照情况。

训练该可变形卷积神经网络时,首先需要数据进行数据集的

标注和扩增。通过拍摄大量仪器的实际图片,分割出其中的字符部分并进行标注。实验采用的汽车尾气分析仪主要有 6 个检测项,对应的标注为:示数 0-9 对应的标签为 0-9,带小数点的数字 0~9 对应的标签为 10-19,负号标记为 20,检测指标的图像按照其名称标记为 21,22,⋯,26,即总类别数为 27。

数据集的扩增通过对保存的图片进行一定角度的旋转、亮度抖动、随机翻转等方式组合完成,以增加训练样本的多样性,并保证每个类别图像的数量平衡。其中,图像旋转的角度在 -15° 到 15° 之间;亮度在进行归一化后进行随机抖动,抖动范围为 ± 0.1 。最后,按照 6:2:2 的比例将数据集分为训练集、验证集和测试集。在训练中,将训练图片归一化为 32×32 像素送入神经网络,对每批图像的相同位置的像素值求平均,计算 mean image 并进行 mean subtraction 操作。

为了得到更好的训练效果及更好地学习目标形变信息,该可变形卷积神经网络的训练分两步进行。首先构建相同结构的普通卷积神经网络并将其训练至收敛,保存模型参数并加载至可变形卷积网络中进行微调,以学习卷积核的形变信息。训练中,以均值为 0、标准差为 0.01 的截断随机正态矩阵初始化神经网络的连接权重;通过 Keras 中的 multi_gpu_model 模块,使用双 Titan Xp GPU 以并行计算的方式进行训练并以单 GPU 方式保存模型参数,设定训练的最大迭代次数 epochs 为 5000,批大小 batch_size 为 32;采用分类交叉熵 categorical_crossentropy 作为损失函数,使用 Adam 优化器实现残差回传算法进行参数更新;设定初始学习率为 0.001,训练中每 5000 步进行学习率衰减,衰减系数为 0.005。

3.2 结果与分析

为了验证本文算法的效果,通过 3.1 节所述方法训练可变形卷积神经网络以及 LeNet5, AlexNet 等经典网络模型,并在划分的测试集上进行测试。各算法的效果如表 2 所列,其中本文提出的可变形卷积神经网络得到的准确率为 99.45%,相比之下,不添加可变形卷积的网络测试的准确率为 97.95%,因此可变形卷积的加入有效提升了模型的识别准确率,使得本文提出的可变形卷积神经网络相对原结构产生了更好的识别效果。

表 2 几种算法在数据集上的效果对比

Table 2 Comparison of several algorithms on dataset

Model	Accuracy / %
LeNet5	97.43
AlexNet	98.15
VGG16	99.07
Our Method	99.45
Our Method without deform-conv	97.95

为进一步验证本文算法的效果和性能,分别对其进行相应的简化测试和干扰测试。

(1) 模型简化测试

简化测试(Ablation Experiment)指去除模型或算法中的部分特性进行对比,以检查该特性对于算法效果的影响。对本文提出的可变形卷积神经网络进行简化测试,即在该模型的基础上,分别添加与删去一个可变形卷积层,比较 3 个网络的性能,以验证可变形卷积的加入对于网络效果的影响。将图

5 中的原始可变形卷积神经网络结构标记为 DCN@conv3~4,将只包含 conv22_offset 单个可变形卷积层的网络记为 DCN@conv4,将在卷积层 conv12 前加入可变形卷积的结构标记为 DCN@conv2~4。使用相同方式训练 3 个网络模型并在测试集上进行测试,对各模型的测试准确率、训练损失函数值以及平均测试速度进行比较,结果如表 3 所列。

表 3 简化测试的结果

Table 3 Results of ablation test

Model	# Deform Layers	Accuracy / %	Loss	Test speed / fps
DCN@conv3~4	2	99.45	0.59	10.8
DCN@conv2~4	3	99.45	0.51	7.5
DCN@conv4	1	98.33	0.93	16.4

简化测试的结果验证了本文提出的可变形卷积神经网络的准确率和时效性。对比实验表明,该网络在保证识别准确率的前提下最大限度地提高了识别速度;同时,可变形卷积层数对识别结果的影响,证明了在卷积神经网络中加入可变形卷积层的有效性和充分性。

(2) 干扰测试

对本文设计的可变形卷积神经网络的识别效果进行分析,识别错误的图像如图 7 所示。可以看出,混淆情况主要发生于 CO 与 CO₂ 之间、NO 与 HC 之间,以及数字 6 与 9 之间。混淆全部发生于扩充数据,且主要存在检测项名称中。

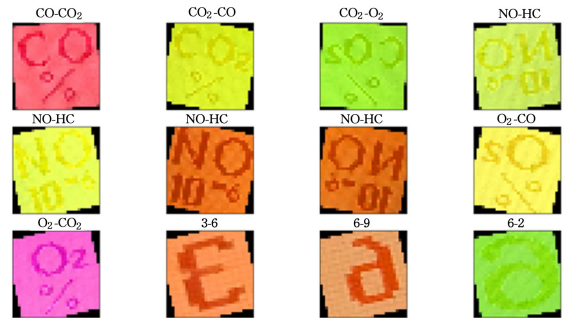


图 7 错误识别图像

Fig. 7 Incorrectly predicted images

为减少模型的混淆情况,提升模型的鲁棒性,对模型进行干扰测试,采用随机方式生成一定的高斯噪声添加到测试图像,并比较其与原图像的识别效果。经测试,该网络在添加噪声的测试集上预测准确率仅降低 0.23%,对测试图像噪声具有一定的抗干扰能力。干扰测试图像如图 8 所示,其中上方为原测试图像,下方为噪声干扰图像。

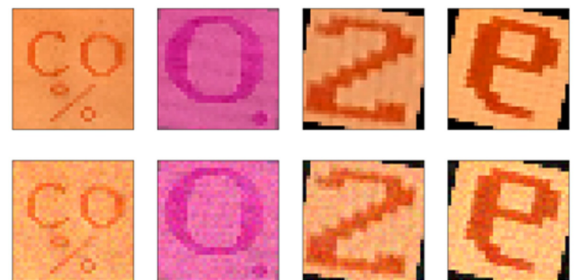


图 8 噪声干扰图像

Fig. 8 Noise interfered image

结束语 为了解决仪器图像识别准确率低以及实际应用中光照等因素干扰的问题,本文基于 GrayWorld 算法及可变形卷积思想,针对汽车尾气分析仪等具有较大屏幕且规则排版的仪器图像,提出了一种图像处理方法与可变形卷积神经网络。在分割中,通过 GrayWorld 色彩均衡再通过彩色分割的方法定位显示屏区域,通过垂直膨胀的方法弥合小数点与对应数字之间的纵向间隔并使用投影分割法进行字符分割,实现了小数点与对应数字的整体识别;使用可变形卷积的思想,对卷积核添加二维偏置,从而更好地学习目标的形态差异,提升模型的准确率。实验结果表明,该网络相比未添加可变形卷积的相同结构的网络准确率提升了 1.5%,达到了较高的识别准确率,且每秒可完整处理 10 张测试图像,能够满足实际应用需求。在实际运用中,可在监测站仪器上加装摄像头,通过本系统进行实时数据识别,也可将算法部署于服务端面向多个设备提供远程检测服务。在后续的研究中,将着重提升图像的分割速度,同时将提高复杂环境下的识别准确率作为重点内容之一。

参 考 文 献

- [1] CUI W X, CUI Y C, WANG Z H, et al. License plate character segmentation algorithm based on template matching and vertical projection[J]. Journal of Qiqihar University (Natural Science Edition), 2015, 31(6): 12-16.
- [2] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International Journal of Computer Vision, 2004, 60(2): 91-110.
- [3] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). San Diego, USA; IEEE, 2002, 1: 886-893.
- [4] DENG J, DONG W, SOCHER R, et al. ImageNet: A Large-Scale Hierarchical Image Database[C]// IEEE Computer Vision and Pattern Recognition (CVPR). 2009.
- [5] RUSSAKOVSKY O, DENG J, et al. ImageNet Large Scale Visual Recognition Challenge[J]. International Journal of Computer Vision (IJCV), 2015, 115: 211-252.
- [6] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
- [7] LECUN Y, CORTES C. MNIST handwritten digit database [EB/OL]. <http://yann.lecun.com/exdb/mnist>, 2010.
- [8] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolution neural network[C]// Advances in Neural Information Processing System. Cambridge: MIT Press, 2012: 1097-1105.
- [9] LECUN Y, BENGIO Y. The Handbook of Brain Theory and Neural Networks[M]. Cambridge: MIT Press, 1998: 255-258.
- [10] XIE S, GIRSHICK R, DOLLAR P, et al. Aggregated Residual Transformations for Deep Neural Networks[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2017.
- [11] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]// ICCV. 2017.
- [12] DAI J, LI Y, HE K, et al. R-fcn: Object detection via region-based fully convolutional networks[C]// NeurIPS. 2016.
- [13] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// CVPR. 2014.
- [14] HE K, GKIOXARI G, DOLLAR P, et al. Mask r-cnn[C]// IC-CV. 2017.
- [15] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Boston, USA, 2015: 1-9.
- [16] LIN M, CHEN Q, YAN S. Network in network[EB/OL]. <http://arxiv.org/abs/1312.4400>, 2013.
- [17] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409.1556, 2014.
- [18] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]// CVPR. 2016.
- [19] BOUREAU Y L, PONCE J, LECUN Y. A theoretical analysis of feature pooling in visual recognition[C]// ICML. 2010.
- [20] FELZENSZWALB P F, GIRSHICK R B, MCALLESTER D, et al. Object detection with discriminatively trained part based models[C]// TPAMI. 2010.
- [21] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks[C]// NIPS. 2015.
- [22] LUO W, LI Y, URTASUN R, et al. Understanding the effective receptive field in deep convolutional neural networks[J]. arXiv: 1701.04128, 2017.
- [23] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]// ICLR. 2016.
- [24] DAI J F, QI H Z, XIONG Y W. Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy; IEEE, 2017: 764-773.
- [25] WANG F, XIANG D. Digital instrument identification method based on convolutional neural network[J]. Machine Design and Manufacturing Engineering, 2018, 9(47): 63-66.



GUO Lan-ying, born in 1963, professor. Her main research interests include intelligent transportation system and so on.



HAN Rui-zhi, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include deep learning and computer vision.