Computer Science and Application 计算机科学与应用, 2022, 12(6), 1631-1640
Published Online June 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.126164

Hans汉斯

基于神经网络的金属零件表面字符检测与识别技术研究

龚　读，景　航，李　聪，聂振康，冯　川，周　浩

四川航天川南火工技术有限公司，四川　泸州

收稿日期：2022年5月24日；录用日期：2022年6月22日；发布日期：2022年6月29日

摘　要

基于神经网络的零件表面字符检测与识别技术研究是通过对产品表面字符特征进行提取，再对提取的特征图像进行分割训练，从字符检测和字符识别两方面进行研究。字符检测方面，首先构建CTPN神经网络模型对原始图像进行特征分析检测，再用贝塞尔曲线进行改进。字符识别方面，首先搭建CRNN神经网络模型对于已处理的特征图像进行识别分析，将特征图像在CNN神经网络中进行特征提取，其次，将已提取的图像输入到RNN神经网络中进行分析预测，最后，再输入到CTC层进行字符转化并输出结果。经过试验测试验证，可以应用于对于产品表面字符识别。

关键词

神经网络，字符识别，OCR检测

Research on Part Surface Character Detection and Recognition Technology Based on Neural Network

Du Gong, Hang Jing, Cong Li, Zhenkang Nie, Chuan Feng, Hao Zhou

Chuan Nan Machinery Plant of China Aerospace Science and Technology Corporation, Luzhou Sichuan

Received: May 24th, 2022; accepted: Jun. 22nd, 2022; published: Jun. 29th, 2022

Abstract

The character detection and recognition technology of parts surface based on neural network is studied by the character feature of product surface. Then the extracted image is segmented and trained, and the character detection and character recognition are studied. In terms of detection,

文章引用: 龚读, 景航, 李聪, 聂振康, 冯川, 周浩. 基于神经网络的金属零件表面字符检测与识别技术研究[J]. 计算机科学与应用, 2022, 12(6): 1631-1640. DOI: 10.12677/csa.2022.126164

firstly, the neural network is constructed to detect the original image and harvest improvements with Bezier curves. In character recognition, the CRNN neural network model is first built to recognize the processed feature images. Then, feature images are extracted in CNN neural network. Secondly, extracted images are input to RNN. Finally, it is input to CTC layer for character conversion and output results. It can be applied to product surface character recognition.

Keywords

Neural Network, Character Recognition, OCR Detection

1. 引言

随着机械加工及装配过程信息追溯需求的增加，越来越多的金属零件在机加环节需要刻印二维码，将关键信息保存在二维码中以实现在接下来的工步中得到该产品在之前环节的关键参数。当前存在的问题在于产品在经过例如特种加工或表面处理等不同的工步后，二维码变模糊从而无法扫描的情况。基于神经网络的零件表面字符检测与识别存在一定的应用可行性，因此针对此项技术进行了验证和研究。

通过对现场需要对产品表面字符信息进行记录的工步进行分析汇总，确定标识字符识别应用的场景，以需要表面字符读取的典型产品为研究对象，梳理所需工步及产品结构，通过对产品在工业镜头下成像、成像图像识别的研究，进行功能模块化设计。基于字符检测与识别的流程，进行标识字符识别项目的功能设计，实现相应产品表面的字符识别。确立了研究内容为以下方面：

1) 工业相机对产品表面的字符聚焦

由于机加产品种类多，大小不一致，对于每一件待识别产品在聚焦层面上，需要对相机捕捉字符对焦机制进行探讨。工业相机本身无法进行变焦，原因在于其成像角度比较低。景深相对较小，现行模式为手工进行调焦处理，对于标识字符的识别将对相机变焦方面进行研究。

2) 针对产品表面文本检测机制

针对对于产品表面的字符区域进行识别，将探讨多种机制以更加精准的方式进行字符区域的捕捉。例如将借助 CTPN 神经网络对于原始图像进行特征分析，或者是 RPN 神经网络进行特征识别。

3) 针对产品识别区域的字符识别机制

将产品表面的字符识别出来以后，针对文本识别，将尝试借助 CRNN 神经网络对于已处理的特征图像进行识别分析和探讨。

2. 零件表面字符检测与识别系统设计

2.1. 系统框架

系统实验环境为操作系统 win10、运行内存 8G、开发工具为 PyCharm、开发语言为 python。实现的 OCR 模型主要使用端到端的方法，通过 CTPN + CRNN 两个神经网络对于文本图像的字符进行了检测和识别。技术路线如下：测试数据输入到 CTPN 神经网络后，经过提取特征信息、BLSTM 融合前后文信息，进行损失函数分析，再输入到 CRNN 神经网络中，通过提取特征信息，预测特征序列，字符转义，将识

别后的字符进行输出。

如图 1 所示，对每一个字符区域进行标记渲染并输出置信度。在将其输入到 CTPN 网络后，CTPN 网络首先会对其进行检测字符的位置并进行标记，再将标记后的这些字符区域图片输入到 CRNN 网络中，CRNN 网络的再对这些字符区域的字符进行识别并输出，通过这两个神经网络的组合，真正实现了端到端的系统，只需要将训练好的网络框架应用到端到端之间，便可以将其用于 OCR 的字符识别，同时保证了有效性和健壮性。

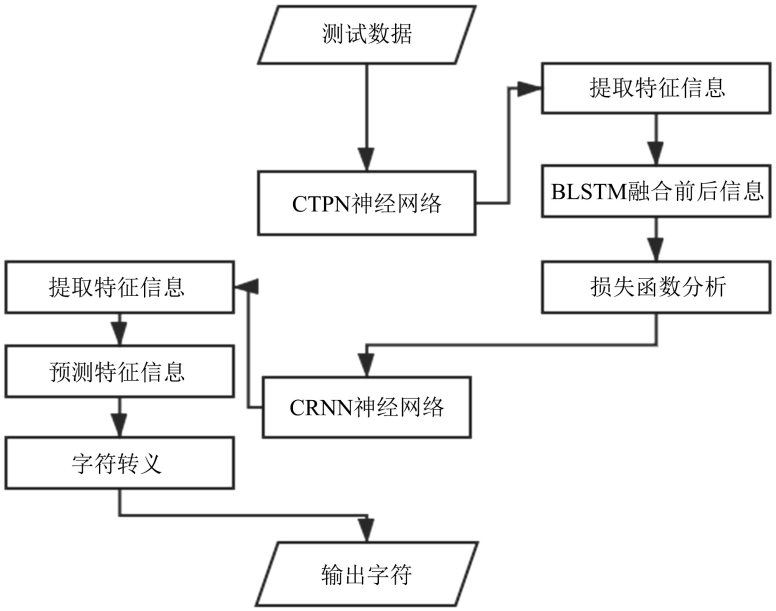

Figure 1. OCR model architecture
图 1. OCR 模型架构

2.2. 数据集

数据集来源于对起爆器现场产品进行数据录入，使用海康卫视工业相机拍照了 1000 份数据，对数据集中字符位置矢量信息和字符信息进行标识记录，再输入到神经网络中进行训练，如图 2 所示。

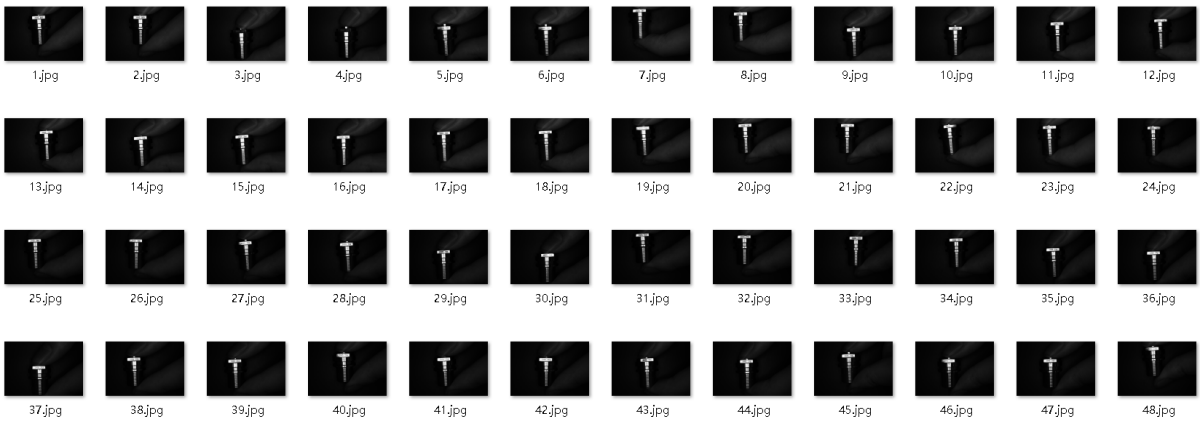

Figure 2. Field collection pictures
图 2. 现场采集图片

3. 系统详细设计

3.1. 检测模块程序设计

对于文本图像的特征提取，经典方法是候选区域探测法，即通过宽高不一的滑动窗口在图像上进行滑动分析，对于潜在的字符的区域进行标记，再将候选字符区域进行归一化处理，将处理后的数据输入到 CNN 神经网络中。

之后，通过卷积神经网络进行层层卷积，池化操作将处理后的数据进行特征提取，根据预期要求来确定固定维度并输出。

最后，将固定维度的输出向量根据特征训练分类器进行分类[1]，得到不同的数据。由于最终检测的实际目标在上述步骤会出现针对一个字符或者多个字符生成多个子区域的情况，所以为了得到精准目标区域，需要借助边界回归(bounding-box regression)来对当前的前景目标进行精确的定位和合并，避免出现重复冗余检测。

针对第二个问题的解决办法：

相比于去预测包含字符的文本框的宽度，预测高度更加可靠和现实一点。因为字符一般是水平排列的，同一水平线上字符的个数未知，而在竖直方向上是可以预知的。除此之外，自上而下的检测方法是先去检测文本所在区域，再去将一个个包含字符的框连接起来，传统的自下而上的检测方法先将单一的字符一个一个通过笔划等匹配识别后，再去合并预测字符区域，没有很好的考虑前后字符区域衔接的问题，不够鲁棒性和健壮性，而且需要较多的子模块去维护，复杂且冗余。

CTPN 的原理基于 Faster-RCNN，首先改进了区域生成网络 RPN [2]，将每一个滑动窗口的宽度进行了限制，将其设定为 3。并增加了长短期循环神经网络(LSTM)，将前面的信息不断地循环操作，来为后面的事件提供预测辅助信息，从而更好地将前后文本信息进行拼接。增加了边界细化，可以使文本框的边界细节更加精准。

CTPN 实现过程主要分为两个部分：检测小尺度文本框，小尺度文本框合并。首先利用 VGG16 提取特征，在特征图上进行滑动窗口检测得到特征向量，之后利用 BLSTM 进行特征向量排序，经全连接层输入到三个不同的输出层中进行输出。第二部分如图 3 所示，通过文本线构造算法来对小尺度文本框进行合并。

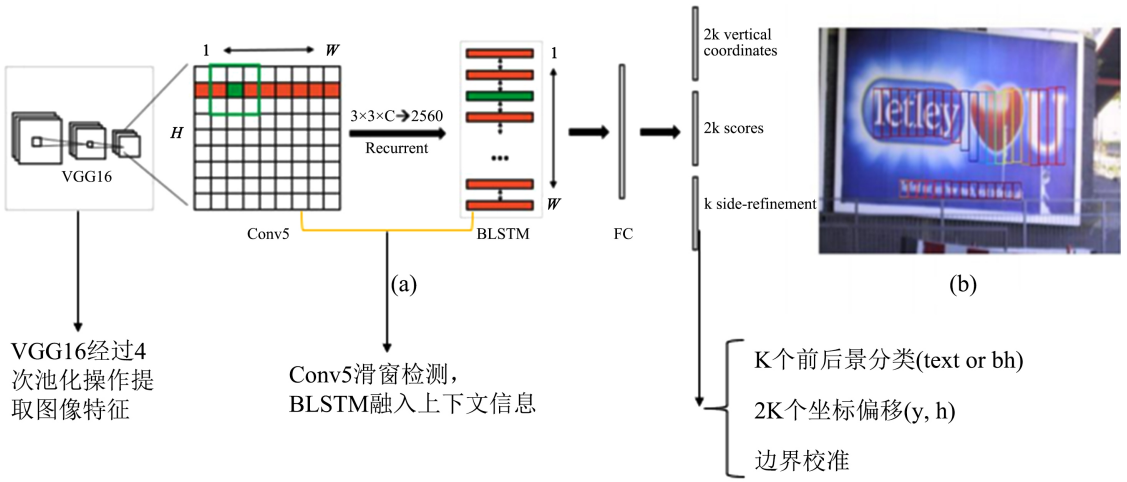

Figure 3. Process of CTPN detecting small-scale text boxes
图 3. CTPN 检测小尺度文本框过程

3.1.1. 提取特征

检测字符的第一步是对原始图像进行特征提取[3]，如图 4，图像文本的特征借助 VGG16 卷积神经网络来提取，VGG16 的 conv5 层将会经过 4 个池化层，层层进行卷积，维度为 16，即输出一副 W × H × C × N 的特征图。经过四次池化后，在 conv5-3 层，特征图的一个像素对应原始输入的 16 像素。

3.1.2. 滑动检测

得到特征图后，划分为 16 × 16 的网格图，进行滑动窗口检测，每一个滑动窗口的大小为 3 × 3，每个窗口检测锚都会包括 9 个小格，每个点都会结合滑窗的区域大小来生成一个新的特征向量，特征向量的长度为 3 × 3 × C。一共设置 10 个检测锚来判定文本字符的范围，用来判定点和锚之间的位移。新的特征图的大小将由 W × H × C × N 变为 N × 9C × H × W。

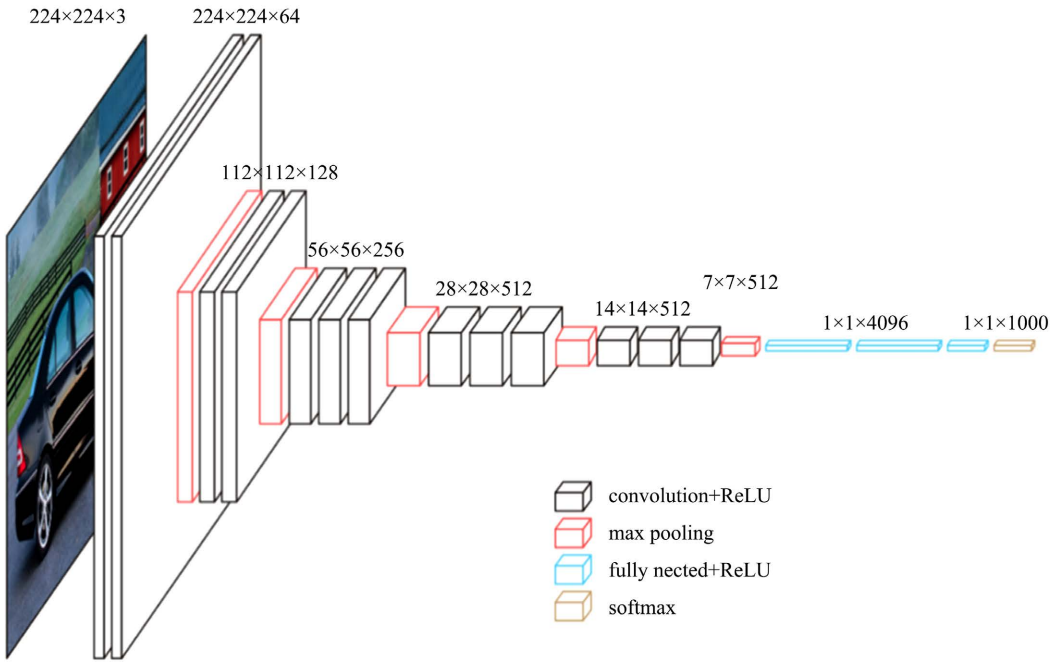

Figure 4. Feature extraction process
图 4. 提取特征过程

CTPN 设置了 10 个等宽度的锚组，每个锚组的宽度一定，高度不一定。锚组的宽度和高度为：Width = [16]，heights = [11, 16, 23, 33, 48, 68, 97, 138, 198, 283]。经过不同的高度设置可以保证每个潜在的字符被锚探测框所覆盖，在被锚探测框所探测选定后，CTPN 首先会进行背景判定，判定锚探测框是否包含文本，其次再用边界框回归分析来修正锚探测框的高度。回归分析公式如下：

$$v_c=\left(c_y-c_y^a\right)/h^a, v_h=\log\left(h/h^a\right), v_c^*=\left(c_y^*-c_y^a\right)/h^a, v_h^*=\log\left(h^*/h^a\right) \tag{2.1}$$

其中，对于背景中字符可能值进行计算的是 $v=\left(v_c^*,v_h^*\right)$，对于锚检测进行高度分析的数值是 v_y^a,h^a，对于边界框线进行预测后的坐标为 $v=\left(v_c,v_h\right)$。

3.1.3. 输出结果以及损失函数分析

FC 层连接有三个分类层，第一个分类层输出 2 k vertical coordinate 来判定矩形文本框的长度，因为宽度锚已确定为 3，所以只输出高度信息，通过判定文本区域的垂直坐标 $V=\left(V_c,V_h\right)$，从而判定字符预测框的高度。第二个分类层输出 1 k anchor 用来判定检测区域是字符还是背景，以 s 用来预测所选图像的文

本/非文本得分，这里通过背景算法设定了一个数值，当大于这个数值就会被判定为字符。因为有些边界框可能会出现背景与字符共存的问题，所以第三个分类层 1 k side-refinement 用来对于边界框进行精确化处理，通过精密优化边框线来使得文本线中的字符比例达到最大。

针对以上三个分类层，CTPN 各设置一个损失函数来对模型进行优化，其总的损失函数为：

$$L\left(s_i,v_j,o_k\right)=\frac{1}{N_s}\sum_i L_s^{cl}\left(s_i,s_i^*\right)+\frac{\lambda_1}{N_v}\sum_j L_v^{re}\left(v_j,v_j^*\right)+\frac{\lambda_2}{N_o}\sum_k L_o^{re}\left(o_k,o_k^*\right) \tag{2.2}$$

判定区域是背景还是字符的损失函数为第一个，其 $L_s^{cl}\left(s_i,s_i^*\right)$ 代表背景中字符的可能值，若是大于 0.5 则会被判定为字符；

判定文本线高度的损失函数为第二个，当第一个损失函数得出为字符的结果后，便可以通过回归分析来判定高度；

对于文本线边界修复的损失函数为第三个[4]，作用是将边界线进行进一步精密。

3.1.4. 文本线构造算法

对于之前所获得的特征图进一步重塑，通过双向记忆神经循环网络(BiLSTM)，每一列的序列特征进行学习，输入到 RNN 网络以后，特征图为 N × 256 × H × W。RNN 将包括循环记忆网络得到的序列特征以及空间特征。经过全连接层，将 RNN 网络最终的序列特征 N × 256 × H × W 输入到全连接层的 512 维中。因为预测的文本探测框将会密集且冗余，针对此问题，可以借助非极大值抑制算法，可以对相近重叠的锚探测框进行清除。核心思想是将包含文本信息的细长的文本探测框进行合并成文本线。

假设在已经选定的图像中已经得到两个文本探测框，各包含一组锚组，且用颜色相互区分，首先对细分的锚探测框进行水平的排序，再通过正反设定的规则得到有连续性的锚组，最终对锚组进行合并。

首先对每一个锚组框，假定为锚组 i 进行正向有限距离查找，当背景值 ＞0.7 的时候可以视为一个字符，选取其后续序列 score 值最大的一个锚组 j；

其次再进行反向有限距离查找，当背景值 ＞0.7 的时候可以视为一个字符，依然选取其后续序列 score 值最大的一个锚组 k；

最后进行对比判定，如果 score 值锚组 i >= score 值锚组 j，则说明是一个最长链接，反之，如果 score 值锚组 i < score 值锚组 j，则说明这不是一个最长的链接组。

下面进行举例：

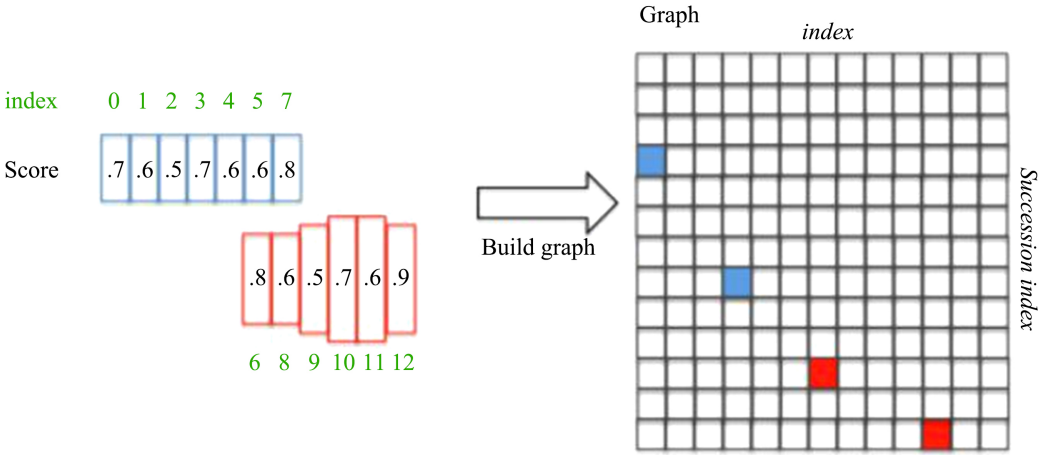

Figure 5. Text detection line process
图 5. 文本检测线的过程

如图 5 所示，锚组已经根据水平线进行排列，并且其背景值都大于 0.7 说明锚组框内都是字符。当 i = 3 的时候，向后寻找 score 值最大的锚组，发现当 i = 7 的时候 score 值最大，在进行反向寻找，发现最大的 score 值是 i = 3，所以 Graph (3, 7)是一组连续的文本探测框。

3.2. 识别模块程序设计

CRNN 的核心思想是认为字符识别本质上是对序列的预测方法，如图 6 所示，CRNN 由三个模块组成，最底层是卷积神经网络，作用是提取特征信息；第二层是循环神经网络，作用是对提取的特征信息进行预测，第三层是字符语义转录，作用是将上一层预测结果转化为字符。

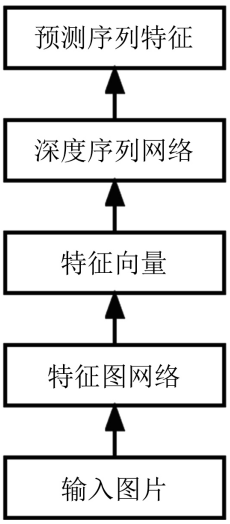

Figure 6. CRNN network architecture
图 6. CRNN 流程

3.2.1. 卷积神经网络层介绍

如图 7 所示，由于 CTPN 神经网络已经将图像中的文本信息进行提取，此时的图片是一个一个包含有字符的小图像。首先对这些图像的大小格式进行统一化处理，之后作为输入数据输入到卷积网络中，卷积网络会对这些图像进行特征分析并进行输出，在池化的最后两次则会将其格式做进一步处理，目的是为了在循环网络中输入数据不再进行处理直接输入。

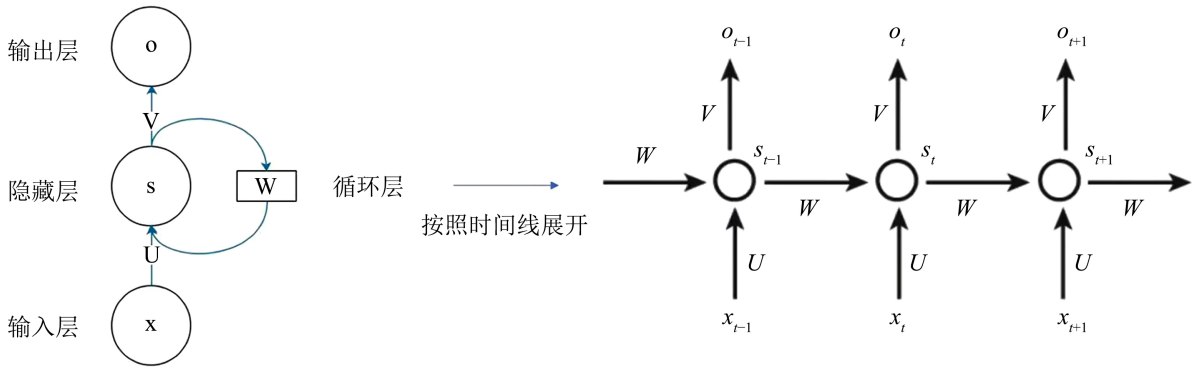

Figure 7. RNN neural network
图 7. RNN 神经网络

3.2.2. 循环神经网络层介绍

循环神经网络的作用是将上一层网络所输入进来的特征序列信息进行预测分析，通过采取双向循环以及 LSTM 神经单元，可以结合前后文信息对当前节点进行更好的排序。

3.2.3. 字符转义层

CTC 的全称是 Connectionist Temporal Classification，直译是对预测进行连接，可以很清晰的明白其作用：对齐输入和输出结果，将 RNN 网络输出的预测信息转换成字符。

关于 CTC 的算法详解：以文本信息举例，已知一个文本字典库和一堆已经预测好的文本信息，接下来的任务是对他们逐一进行匹配，可以先定义规则，对于不同的字符可以根据他们的字符组成坐标位置来判定是哪些字，同样也可以应用到音频的检测中，但很明显也存在一些问题，无法得知输入和输出是否匹配。CTC 假设标签之间是独立的，并且每次输出都是单个字符的概率，利用贝叶斯定理计算预测序列的后验概率分布，所以只是针对局部信息进行预测。可以用序列 X = [x1, x2, ···, xt]表示输入信息，用 Y = [y1, y2, ···, yu]表示输出信息，只需找出它们之间的映射关系，在 RNN 网络输出预测信息后，假设输出了 M 条预测信息，说明在字典库中将会有 M 个字符进行输出。但是预测并不是 100%准确，每条预测信息都是由概率的，通过极大值可能性去分析每一条预测信息，将极大值结果进行输出，匹配率一般比较高。

3.3. 相机传输模块设计

相机传输模块主要分为两部分：

第一部分为对产品表面的字符成像进行实时传输到检测网络中，通过检测网络对字符区域进行检测，再将检测结果传输到识别网络中，通过识别网络将识别结果输出。工业相机在试验验证阶段采用海康威视工业相机，海康威视工业相机具有 + 语言版本的 SDK 开发包，可对相机进行二次开发。在项目中，对于工业相机主要开发的层面在于使相机可以将图像通过串口通讯等手段将图像实时传输到检测网络中。因此，将开发针对 SDK 开发包与 python 识别程序之间的串口之间的通讯协议。

第二部分为产品表面成像，刻在产品壳体表面的字符，会因为光源导致部分字符在相机镜头下成像效果由于反光不清晰。工业镜头常用的光源为环形光源，经过调研，计划采用栅形光源，字符的成像将更为清晰。

4. 检测结果

4.1. 检测率分析

每张文本信息均有分布在好几处的文本信息，因此每张图像上所能检测到文本框数目也不同，在这里通过将每个类别的图像的检测率都找一个代表来展示一个较为直观的效果。

对于以上四种类型的文本信息图片，分别进行检测，再进行检测率分析，通过图 8 以看出，对于包含文本信息的图片在通过滑窗检测后基本上都能找出字符所在的位置并进行标记，并且检测到的准确率能达到 90%以上。

基本上对于字符都可以进行标记并检测出来，对于名片类截图类的文本信息检测的精准性高一些，但即使对于背景噪点大，文本信息不突出的街景类信息检测的精准性也很高。例如对于生活类广告牌，如果文字以水平方式排列，则会 100%检测出，如果背景噪点导致像素模糊不清晰的情况下，可能无法判定是否为字符或者背景。

图 8 通过五组图片测试集，包含 81 个合并文本线检测框，可以对于检测准确率进行直观的展现。

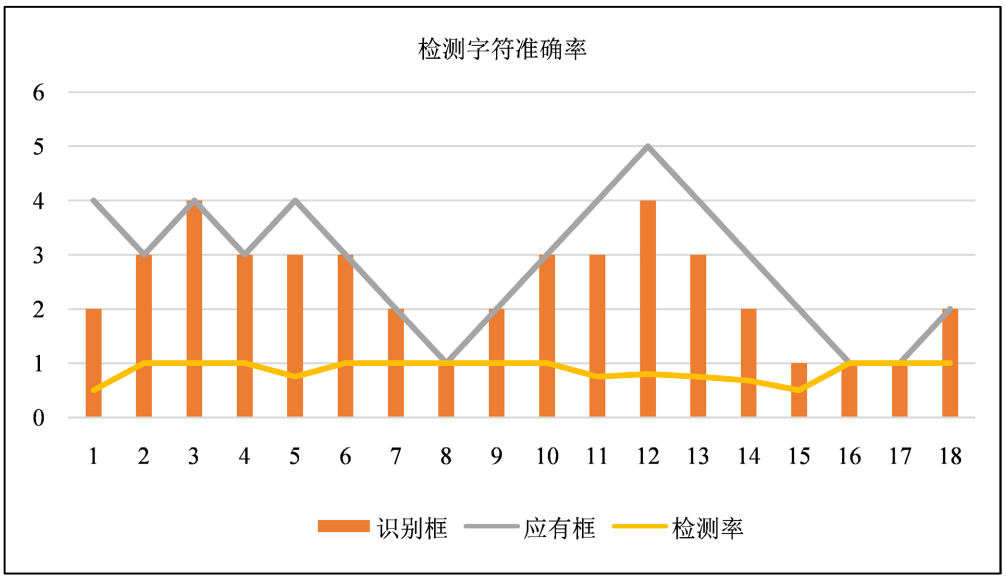

Figure 8. Character accuracy detection
图 8. 检测字符准确率

4.2. 识别率分析

针对每个大类中的代表性图像进行输出文字分析。其中，由于每个图像中的文本信息不同，所以包含的识别框不同，首先对于其真正包含字数进行分析，再将神经网络中识别后的字数进行统计，结果如图 9：

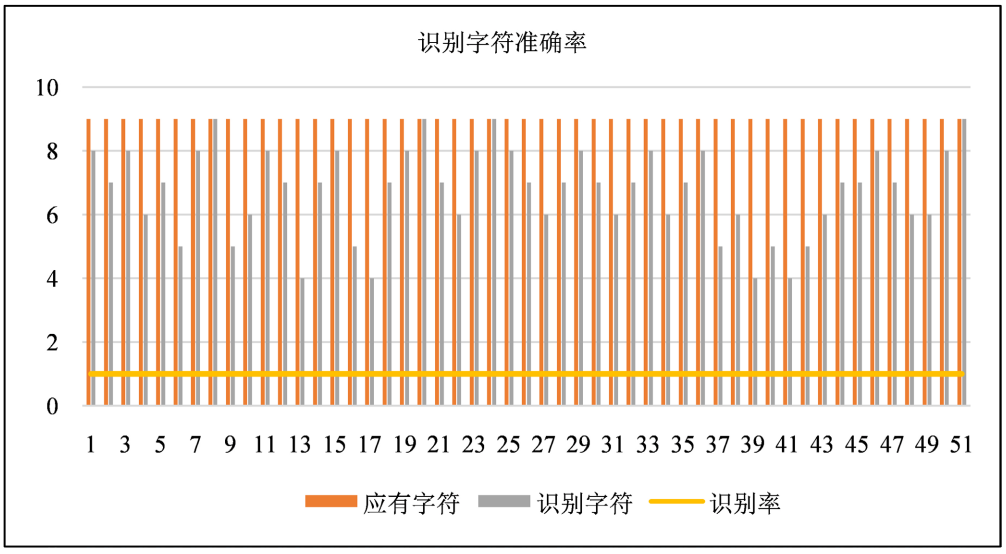

Figure 9. Character recognition accuracy
图 9. 识别字符准确率

通过以上图像可以看出，对于名片类，截图类图像文本，识别率几乎可以达到 100%，然而对于背景噪点比较大的自然场景文本，尤其时文本弯曲程度比较大的时候，其识别率将降低，最低可以降到 65%，这也可以看出对于神经网络仍有很大改进的地方。

4.3. 数据综合分析

通过对于检测率和识别率的展示，可以看到检测率均能达到95%以上，对于识别率平均也有89%，因此整体上OCR对于场景文本识别尤其是排列整齐的文本检测识别效果非常好，但依然存在缺陷，就是在成像为弯曲的时候，识别率会大大降低。

5. 结束语

基于神经网络的零件表面字符检测与识别技术研究，达到了以下效果：

1) 形成了一套关于产品表面字符识别软件，对于训练数据集后的单发产品的检测率可达到90%以上；

2) 标识字符识别已进行实例测试，为后续实现现场的字符识别提供了技术基础，同时该技术可以推广应用于各类型号的产品表面字符识别。

参考文献

[1] 蒋良卫, 黄玉柱, 邓芙蓉. 基于深度学习技术的图片文字提取技术的研究[J]. 信息系统工程, 2020(3): 87-88.

[2] 杨捷, 刘进锋. 利用CTPN检测电影海报中的文本信息[J]. 电脑知识与技术, 2018, 14(25): 213-215. https://doi.org/10.14004/j.cnki.ckt.2018.3104

[3] 刘基. 基于OCR技术的低对比度下铸件标识字符检测与识别研究[D]: [硕士学位论文]. 太原: 太原科技大学, 2021. https://doi.org/10.27721/d.cnki.gyzjc.2021.000079

[4] 邵慧敏, 张太红. 基于CTPN神经网络对营业执照文字检测模型[J]. 计算机技术与发展, 2021, 31(1): 94-97.