

DOI:10.19651/j.cnki.emt.2210303

基于改进YOLOv5与CRNN的电表示数识别*

黄辉 肖豪 王琼瑶 吴建强 梁志龙

(五邑大学智能制造学部 江门 529030)

摘要: 为了提高电表示数检测和识别的准确率,基于轻量高效的YOLOv5s网络提出了改进的目标检测网络。首先,在特征提取阶段添加CBAM注意力机制对图像的重要特征进行自主学习,并设计了一种特征融合网络D-BiFPN加强了对深层特征的提取;其次,引入CIOU损失函数,使目标框的回归更加稳定。对CRNN文本识别算法的主干网络进行改进,模型保持轻量化的特点,在移动端部署上有良好的前景。最后,在电表数据集上测试得出:相比于YOLOv5算法,所提出的算法精度均值提升了5.13%;相比于CRNN算法,所提出的文本识别算法准确率提升了7.4%。实验结果表明,改进后的文本检测算法对电表示数的检测精度较高,文本识别算法准确率和速度较高,满足电表示数检测识别的实际应用需求。

关键词: 电表示数识别;YOLOv5;CRNN;注意力机制;文本检测

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 460

Electric meter indication recognition based on improved YoLov5 and CRNN

Huang Hui Xiao Hao Wang Qiongyao Wu Jianqiang Liang Zhilong
(Intelligent Manufacturing Department of Wuyi University, Jiangmen 529030, China)

Abstract: In order to improve the accuracy of meter reading detection and recognition, an improved target detection network is proposed based on the lightweight and efficient YOLOv5s network. Firstly, in the feature extraction stage, CBAM attention mechanism is added to learn the important features of the image, and a feature fusion network D-BiFPN is designed to enhance the extraction of deep features; Secondly, the CIOU loss function is introduced to make the regression of the target box more stable. The backbone network of CRNN text recognition algorithm is improved. The model maintains the characteristics of lightweight, and has a good prospect in the deployment of mobile terminals. Finally, the test on the meter data set shows that compared with the original YOLOv5 algorithm, the average accuracy of the proposed algorithm is improved by 5.13%; Compared with the original CRNN algorithm, the accuracy of the proposed text recognition algorithm is improved by 7.4%. The experimental results show that the proposed text detection algorithm has high detection accuracy and stability, and the text recognition algorithm has high accuracy and speed, which can meet the application requirements of meter reading recognition.

Keywords: electric meter indication recognition; YOLO v5; CRNN; attention mechanism; text detection

0 引言

在当今,电力系统和人们的日常生活和工业生产息息相关,而智能电表应用在电力系统的各个方面。智能电表的产量和使用量迅速增长,在智能电表的自动化检定和电力系统巡检的过程中,电表读数不准确和读取速度慢是必须要解决的问题。一些款式较老的电表不具有后台采集信息的功能,检定时靠人工读数。对于具有后台数据采集功能的智能电表,检定的时候需要检查电表示数和后台收到

的数据是否一致。传统人工读数的方法费时费力,并且读数准确率较低,人工抄录示数时容易出现错误,影响电表检定结果的准确性和客观性。在电表的自动化检定和电力系统的巡检时,急需一种效率高、准确率高的智能识别算法,摆脱传统人工的种种缺陷限制。目前一些团队做了有关电表示数识别的工作,欧家祥等^[1]提出了一种MASK-LSTM-CNN算法来检测电表示数,比传统的Mask R-CNN算法和Faster R-CNN算法的识别精度提高了2%~9%,结果表明有效解决背景信息干扰较多的问题;Zhou等^[2]开

收稿日期:2022-06-13

* 基金项目:国家自然科学基金青年基金(51905384)项目资助

发出了 EAST 算法,利用 FCN 模型在样本目标上回归字符区域的轮廓,结果表明所有像素位置都输出了密集文字预测;Lin 等^[5]提出了一种改进的支持向量机和笔画宽度算法分别对文本进行检测和识别,结果表明可以有效克服字体大小不一、字符倾斜等因素对识别精度的影响;Zhan 等^[6]提出了一种有效识别透视变换过的文本的 RARE 算法,但是该方法收敛速度慢;Belharbi 等^[5]提出了 CRF 与卷积神经网络结合的网络模型,文本检测后使用 recursive RNN 模型,此模型插入了注意力机制,直接利用图像完成字符串学习,然后进行识别;Yang 等^[6]提出一种全卷积序列识别网络对数字进行识别,设计增广损失函数处理数字间隙,减少了计算量和模型参数,但是准确率不够高;Tian 等^[7]提出了 CPTN 算法,将 CNN 与 RNN 网络相结合,同时融合空间和序列特征,但是该算法不能处理复杂文本。

尽管现有的研究已经将电表示数的目标检测和识别与深度学习方法结合,但是这一研究领域仍然存在以下问题:

1) 现有模型训练和模型预测运算复杂、耗时长,无法满足实际的应用需求。

2) 电表示数检测识别任务中,难点在于示数区域在电表图像中的区域较小,不易被完整检测出来。示数中存在小数点,应用传统文本识别算法,难以解决。

针对以上问题,本文提出一种电表示数检测和识别算法,将改进后的 YOLO v5 作为电表示数的目标检测算法,在特征提取网络 CSPDarknet53 中插入通道和空间维度注意力机制并且提出一种 D-BiFPN 替换原网络中的 PANet,使网络可以自主学习输入图像的重要特征并且加强对图像更深层信息的特征提取。其次,引入 CIoU^[8] 损失函数使预测框的回归效果更好,使目标检测的准确率和运算时间都优于原算法。对检测到的文本区域,通过改进后的 CRNN(convolutional recurrent neural network)算法进行文本识别,CRNN 算法无需进行文本分割,对小数点在内的数字文本识别效果准确率高,识别速度快。算法本身不受电表背光颜色、场景、字体大小不一和曝光率等问题的约

束,对各种环境下采集到的图像均有较好的识别能力。

1 基于 YOLOv5 算法的电表示数检测

1.1 YOLOv5 网络架构

本文使用的是 YOLOv5s 模型。YOLOv5s 网络主要包括输入端、Backbone 模块、Neck 模块以及 Head 模块。YOLOv5s 网络的输入端增加了 Mosaic 数据增强算法,用在模型训练阶段,提高了小目标的检测效果。训练数据集时可通过自适应计算得到最佳锚框值,这有利于提升检测速度。

Backbone 模块主要使用 CSPDarkNet53 和空间金字塔池化结构^[9]。Backbone 模块引入一种新的 Focus 结构,它对图像进行切片,把图片的宽度 w 及高度 h 信息整合到通道 c 维度,所以宽度 w 及高度 h 缩小成为原来的 $1/2$,在通道方向上拼接,通道 c 则会增加 4 倍。Focus 模块有效减少了信息丢失、提高了检测速度。空间金字塔池化结构中使用 3 个大小不同的池化核将前一层网络输出的特征图进行池化操作,之后把得到的结果在通道维度上进行连接。YOLOv5s 模型中添加空间金字塔池化结构对推理速度的影响很小,但是增加了特征图的感受野,锚框和特征层对齐的问题就易于解决,使检测效果明显提升。

Neck 模块组合图像特征,传递到预测层,采用由特征金字塔结构^[10] (feature pyramid networks, FPN) 与路径聚合网络结构^[11] (path aggregation network, PAN) 组合而成的 FPN+PAN 网络,增强了网络对缩放尺度不同的特征融合能力。FPN 网络使用上采样操作获得高层特征信息,然后高层信息与底层特征信息融合,增强网络的学习能力,缺点是会丢失一些定位信息。PAN 对 FPN 获得的图像增加了由下向上的路径,使底层定位信息更好传递至顶层,二者一起使用可达到互补的效果,网络特征融合能力有所增强,提高模型的准确性与鲁棒性。其结构如图 1 所示,输出端使用 GIOU 边界框损失函数,在筛选目标框的过程中采用非极大值抑算法,提升多目标和小目标的检测能力。

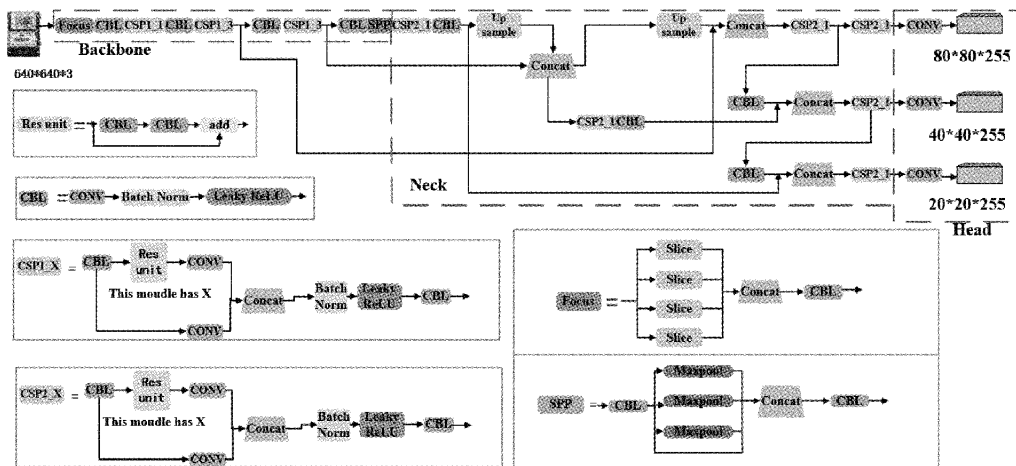


图 1 YOLOv5 网络结构示意图

1.2 改进的 YOLOv5s 结构

本文改进了 YOLOv5s 的 3 个方面:1)插入 CBAM 注意力模块,目的是增加主干网络的特征提取能力;2)提出 D-BiFPN 多尺度特征融合网络;3)改进损失函数。

1) 主干网络的改进

在目标检测任务中,一张图片上不同位置的信息重要程度不同,不同通道的信息重要程度也可能不同。所以,为了提高主干网络对电表图像中示数区域特征的提取能力,引入了注意力机制,该机制是在机器学习模型中嵌入的一种特殊结构,用来自动学习和计算输入数据对输出数据的贡献大小,提升网络对特征的利用率。由于电表示数区域所在液晶显示屏中的区域目标较小,使目标检测的准确率降低。因此,在 YOLOv5 中引入注意力机制。

广泛应用的注意力机制包括 SENET^[12](squeeze and excitation net)和 CBAM^[13](convolutional block attention module),SENET 是通道注意力模块,而 CBAM 是结合了空间与通道两种注意力机制的模块。本文设计了这两种注意力机制的对比实验。

为了减少网络的运算量,本文仅仅在 CSPDarknet53 特征提取网络中插入 CBAM 模块,该模块能在通道维度和空间维度分别计算出注意力的权重,使网络可以更加关注图像中的待检测区域,增强主干网络特征提取能力。通道注意力模块与空间注意力模块串联组合而成 CBAM 模块,其结构如图 2 所示。

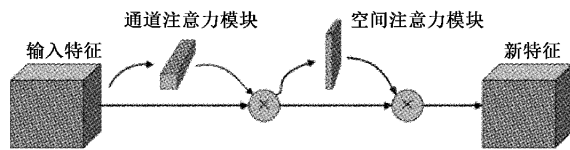


图 2 CBAM 模块示意图

通道注意力机制关注的是哪些位置是有意义的。对于通道注意力模块,特征的每个通道都表示一个专门的特征检测器,将输入的特征进行全局最大池化与全局平均池化,然后对特征图在空间维度进行压缩,经过多层感知器建立模型相关性,之后通过 Sigmoid 激活函数获得权重^[14]。其结构如图 3 所示,公式如式(1):

$$M_c(F) = \sigma[W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))] \quad (1)$$

式中: σ 为 Sigmoid 激活函数; F_{avg}^c 为全局平均池化; F_{max}^c 为全局最大池化; W_0 为共享全连接层的第一层; W_1 为共享全连接层的第二层; c 为输出向量的长度。

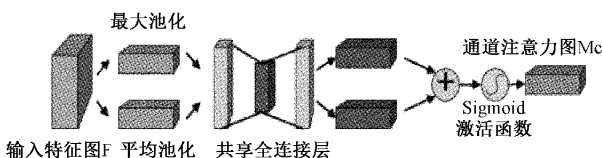


图 3 通道注意力模块

空间注意力机制关注的是哪些特征是有意义的。空

间注意力模块将通道注意力模块的输出作为输入,首先把输入特征图在通道维度分别进行全局最大池化与全局平均池化,然后合并池化后的 2 个通道特征图,使用卷积操作生成空间注意力特征图^[15]。其结构如图 4 所示,公式如式(2):

$$M_s(F) = \sigma(f^{7 \times 7}([\text{AvgPool}(F); \text{MaxPool}(F)])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \in R^{H \times W} \quad (2)$$

式中: $f^{7 \times 7}$ 表示对图像进行 7X7 的卷积; F_{avg}^s 表示全局平均池化; F_{max}^s 表示全局最大池化; $M_s(F)$ 表示输出的特征图。

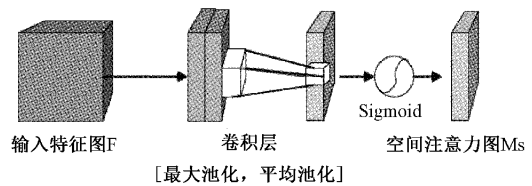


图 4 空间注意力模块

CBAM 模块实现了对输入特征的筛选,重点关注了对模型贡献比较大的特征,忽略用处较小或无用特征,提升了检测的效率,增加了主干网络的特征提取能力。这种模块即插即用,在不显著增加计算量和参数量的情况下可以提升模型的准确率,增强模型的泛化能力。

本文在 YOLOv5 算法中添加 CBAM 模块的位置如图 5 所示。把注意力模块插入到电表示数的目标检测任务中可以自适应地突出体现目标特征信息、增加感受野和削弱无关背景的干扰。

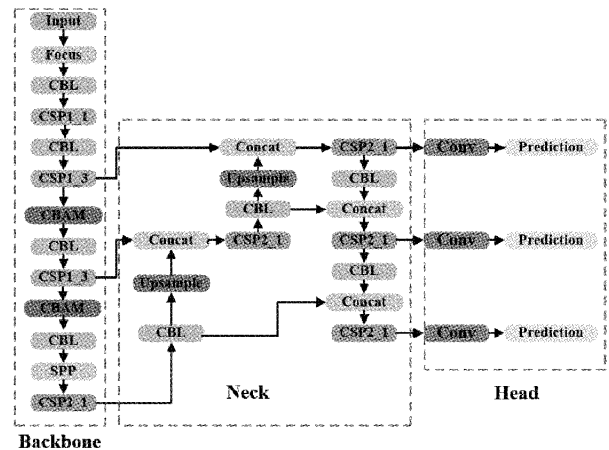


图 5 添加 CBAM 的 YOLOv5 网络

2) 特征金字塔模块的改进

原始的 YOLOv5 网络框架采用的是 FPN + PAN (PANet) 结构进行多尺度特征融合。该网络的融合方式为简单地通过把特征图变换成相同尺寸,然后进行特征图之间的级联,只考虑了上层特征对检测结果的影响,而没有考虑下层特征对检测结果的影响,无法充分利用不同尺度之间的特征,造成网络的检测精度受限,其结构如

图 6(a)所示。

加权双向特征金字塔网络^[16] (bidirectional feature Pyramid network, BiFPN)是在 PANet 基础上改进的,其结构如图 6(b)所示。P3~P7 是 5 个输入特征,经过上采样、下采样及叠加等运算过程,最后输出 5 个提取后的特征。不同输入特征分辨率也有所不同,而且对各级特征融合后的特征贡献程度明显不同,所以 BiFPN 网络为每个输入特征添加了可学习的权重,使网络不断地调整权重以确定每个输入特征对输出特征的重要程度。BiFPN 使用的是快速归一化融合,直接使用权值除以所有权值的加和进行归一化,同时把权值归一化到[0,1]之间,提高了计算速度,如式(3):

$$O = \sum_i \frac{w_i}{\epsilon + \sum_j w_j} \cdot I_i \quad (3)$$

式中: O 和 I_i 表示的是输入和输出特征;权重 w_i 使用 ReLU 激活函数保证 $w_i \geq 0$; $\epsilon = 0.0001$ 是保证数值稳定的小量。

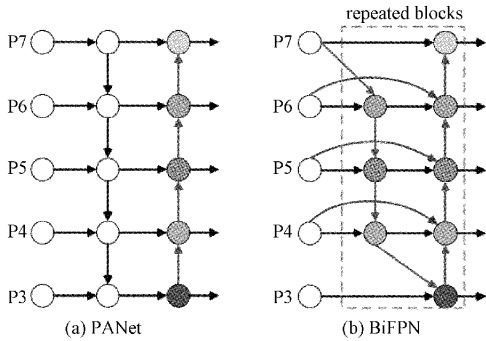


图 6 PANet 与 BiFPN 网络示意图

BiFPN 相比于 PANet,主要改进了三点:

- (1)增加了跳跃连接,输出层不仅可以获得特征融合后的自下而上的信息;
- (2)形成了融合模块,可作为整体继续参与堆叠,做进一步的融合;
- (3)减少了部分结点,删除了入度为 1 的节点,因此相较于前一节点没有额外信息,减少了冗余计算。

虽然 BiFPN 网络加权融合了上下层的特征作为最终的输出结果,但是实际图像采集过程中拍摄设备距离电表远近、角度和光照不同,都会导致部分待检测图像只有少数有用的特征供网络训练学习。

为提高模型的文本检测效率及融合效果,本文在 PANet 与 BiFPN 的基础上进行改进并提出了 D-BiFPN 网络,考虑到模型参数数量和复杂度的问题,所以在原 BiFPN 网络的输出结果后增加了一个卷积层来提取更深层次的特征信息,使网络对小目标物体与有用特征量较少的图像检测能力提升。同时,对进一步提取到的相邻两个特征信息之间进行 Concat 融合操作,提升相邻特征之间的关系,

使多尺度的特征信息被有效利用,从而提高检测电表示数场景下的小目标检测精度。改进后的 BiFPN 结构如图 7 所示。

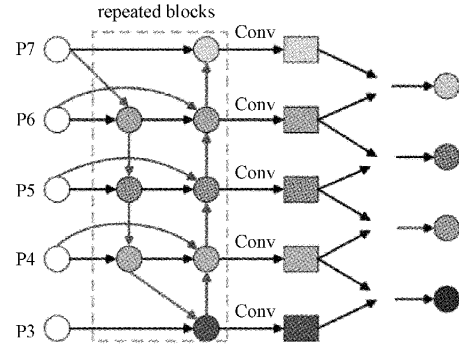


图 7 改进后的 BiFPN 网络

3) 损失函数的改进

YOLOv5 网络的损失函数包括定位损失 (localization loss)、分类损失 (classification loss) 和目标置信度损失 (confidence loss)。YOLOv5 使用 BCELogits 损失函数计算 objectness score 的损失, class probability score 采用了交叉熵损失函数 (BCELoss), bounding box 采用了 GIOU Loss^[17-18] 来计算 bounding box 的 Loss,公式如式(4):

$$L_{GIOU} = 1 - IoU + \frac{|C - B \cup B^{gt}|}{|C|} \quad (4)$$

式中: $B = \{x, y, w, h\}$ 是预测框的大小; $B^{gt} = \{x_{gt}, y_{gt}, w_{gt}, h_{gt}\}$ 为真实框大小; C 为 B 和 B^{gt} 的最小区域的面积; C 与真实框与预测框的距离有关,距离越大则 C 的值越大, $|C - B \cup B^{gt}|$ 也越大,最后趋近 1。

尽管 GIOU 解决了 IOU 的梯度无法有效反向传播更新和边界框不重合时的问题,但是 GIOU 仍然存在边界框不够精细、收敛速度慢与不稳定等问题。因此本文使用 CIOU 作为边界框回归的损失函数, CIOU 损失函数如式(5)所示。

$$L_{CIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (5)$$

式中: c 是两个边界框最小包围区的对角线长度; b 和 b^{gt} 表示的是 B 和 B^{gt} 的中心点; ρ 表示的是预测框中心点和目标框中心点的欧氏距离; α 是一个正的权重参数,不参与梯度计算; v 是衡量长宽比一致的函数,公式如下:

$$IOU = \frac{b \cap b^{gt}}{b \cup b^{gt}} \quad (6)$$

$$\alpha = \frac{v}{(1 - IoU) + v} \quad (7)$$

$$v = \frac{4}{\pi^2} \left(\left(\arctan \frac{w^{gt}}{h^{gt}} \right) - \arctan \frac{w}{h} \right)^2 \quad (8)$$

式中: $\frac{w}{h}$ 与 $\frac{w^{gt}}{h^{gt}}$ 表示的是预测框与目标框之间的宽高比, CIOU 增加了检测框尺度的 loss,增加了长和宽的 loss,这

样预测框与真实框的差异就会更小,使预测框的回归效果更好。

2 基于 CRNN 的电表示数识别

本文尝试了通过 Tesseract 进行电表示数的识别,但是 Tesseract 是基于模板的文本识别算法,对于数显屏幕上的文本识别错误率比较高,特别是对带有小数点的数字识别准确率很低。通过一系列实验,最终选择 CRNN^[19]算法对电表示数进行识别,其对于图像的序列识别具有较好的效果,特别是场景文字识别。

通过 CRNN 算法进行电表示数的识别具有以下优点:1)没有字符长度的限制,只需要输入文本图片和对应的标签序列;2)通过 CNN^[20-22]层和 LSTM 可以直接根据输入图片得到标签序列,无需字符分割;3)模型参数量小,易于部署在移动终端。很多传统算法很难解决小数点的识别问题,在文本分割的时候出现小数点的分割错误,而 CRNN 是一种不需要文本分割的算法,通过特征序列的分类概率来确定真实字符。其分为 3 个部分:卷积层、递归层和翻译层。网络结构如图 8 所示。

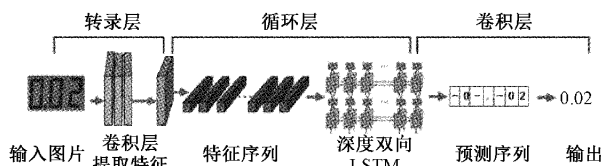


图 8 CRNN 网络结构

CNN 层提取输入图像的特征序列,得到特征图;RNN 层使用双向 RNN(BLSTM)对特征序列进行预测,学习序列中的每个特征向量,并且输出预测真实的标签分布;转录层使用连接时序分类^[23](connectionist temporal classification, CTC)损失,将从 RNN 层获取的标签分布转换成最终的标签序列。

2.1 改进 CNN 卷积层

在 CRNN 原始代码中特征提取部分使用的是 VGG16 网络,但是在特征复用和特征传递方面, DenseNet^[24] 优于 VGG16 网络。DenseNet 是在特征提取方面性能比较突出卷积神经网络模型,通过建立网络中某一层与前面所有层的密集连接使特征被充分利用,并且使用的计算量和参数量更少,进而提升了特征提取的效果。因此,本文的电表示数文本识别用 DenseNet 来替换 VGG16 网络,其网络结构如图 9 所示。

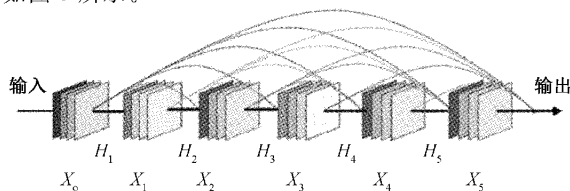


图 9 DenseNet 网络结构图

DenseNet 的核心设计思想是通过密集连接机制把所有网络层相互连接,即将之前每个层的输出都作为当前层的附加输入。神经网络第 l 层的输出 x_l 是对 $l-1$ 层的输出 x_{l-1} 施加非线性函数变换得到的,如式(9):

$$x_l = H_l(X_{l-1}) + x_{l-1} \quad (9)$$

在 DenseNet 网络中,第 l 层与前面所有的网络层在通道维度合并连接,然后作为第 $l+1$ 层的输入。一个 N 层的网络会产生 $N(N+1)/2$ 个连接,第一层的特征与后面四层的特征一起叠加到网络的最后一层进行分析,所有层的权重分散在同一块内的多个输入上,因此网络就可以提取出更高级的特征,这就是特征复用。如果使用 $\{x_0, x_1, \dots, x_{l-1}\}$ 表示把网络的第 0 层到第 $l-1$ 层的输出特征图进行通道合并,那么第 l 层的输出可由式(10)得到,这就是密集连接。

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}]) \quad (10)$$

密集连接方式在密集模块中集中体现,每一个密集模块都包括了批量归一化层、激活函数、池化层和卷积层,这些层相互连接并且特征图的大小相同。两个相邻的密集模块之间通过过渡层来改变特征图的大小,过渡层包括瓶颈层与池化层。其中,瓶颈层是 1×1 的卷积层,作用是降低参数量和输出特征图的数目。这种密集连接方式相当于将每一层都直接和网络输入与损失函数相连接,因此可以在更深的网络结构中,有效解决在训练过程中出现梯度消失的问题。由于 DenseNet 独特的结构设计,某一层网络连接来自其前面所有层的特征图,对提取到的特征实现重复利用,可以使最后损失函数的分类使用低级特性,计算更加简便,目标优化更加高效。

2.2 RNN 层

将 DenseNet 网络提取出的特征序列送入 RNN 网络中进行文本预测,由于 RNN 网络训练过程中会出现梯度消失和梯度爆炸等问题,不能获取更多的上下文信息,因此 CRNN 使用 BLSTM 网络,其特殊结构设计允许它能够在多次训练过程中保持稳定,从而能够获得长距离的上下文信息。BLSTM 单元的内部结构包括存储单元、输入单元、输出单元和遗忘门。存储单元只能存储过去的上下文信息,所以需要 BLSTM 进行双向上下文存储,遗忘门可以清除单元的内存。

2.3 CTC 层

连接时序分类也叫连结主义时间分类是用来解决输入序列与输出序列难以一一对应的难题。CTC 层把循环层输出的每帧预测结果概率转换成标签序列输出,使输入序列和输出序列对齐。预测结果转录而生成的文本会存在大量重复堆叠的现象,简单处理会造成文本信息丢失。例如电表示数 2200.03 经过 RNN 预测的结构可能为 '22200000.003',简单去除重复部分则变为 '220.03',这样简单处理会造成信息丢失。而 CTC 提出了 blank 机制来解决该问题,即通过把某些位置设置为占位符 ' ',在文

本序列中的重复字符之间插入一个占位符,上述的预测结构可能变成‘2-200-00.003’,将没有经过占位符的重复数字去除后就得到了‘2200.03’,这个过程称为解码,而编码则是通过神经网络来实现。在数学模型上,CTC层可根据输入的每帧预测信息寻找具有最高概率分布的标签序列输出。CTC的损失函数定义为:

$$L_S = -\ln \prod_{(x,z) \in S} P(z|x) = -\sum_{(x,z) \in S} \ln P(z|x) \quad (11)$$

3 实验结果分析

3.1 电表示数目标检测实验结果分析

1) 实验环境

本实验所用计算机处理器是 i7-9700,显卡型号为 NVIDIA GeForce GTX 1080 Ti,内存为 11 GB。操作系统为 Windows 10,训练框架选择 Pytorch,软件编程环境为 Python 3.7。

2) 网络训练

本次训练的数据集电表图像有 5 000 张,数据集按照 8:1:1 的比例分成训练集、验证集和测试集。

选用 YOLOv5s 作为目标检测网络,此算法参数量仅为 7.5 M,易于部署在移动终端。主要的训练参数配置如下:输入图片尺寸为 640×640,设置标签平滑为 0.01,防止过拟合。模型训练共设置 600 个 epoch,前 200 个 epoch 采用冻结训练,加快训练速度,防止训练初期破坏权重。初始学习率设置为 0.001,使收敛速度加快。第 201 个 epoch 解冻训练,同时学习率减小为 0.000 1,以更新整个网络的参数权重。批处理量(batch size)为 8。

3) 算法改进前后性能对比

电表示数识别一般应用在变电站巡检和电表的自动化检定,对测试样本计算精度(precision)、召回率(recall),使用精度均值(AP)和检测一张图像消耗的时间作为评估指标。precision 是评价算法的定位和目标检测能力;recall 是判断检测出对象能力的指标。以 recall 为横轴,precision 为纵轴绘制 PR 曲线,对其积分求出曲线下的面积,表达式为:

$$precision = \frac{TP}{TP+FP} \quad (12)$$

$$recall = \frac{TP}{TP+FN} \quad (13)$$

$$A_p = \int_0^1 p(r) dr \quad (14)$$

式中:TP 为分类正确的正样本;FP 为分类错误的负样本;FN 为分类错误的正样本; $p(r)$ 为 PR 曲线函数的表达式。

表 1 所示为在 YOLOv5 模块中添加了不同模块对网络性能的影响。相比于原模型,V1 在 Backbone 中添加了 SE 通道注意力模块,AP 提升了 0.25%, t 缩短 0.004 s。

V2 在 Backbone 中添加了 CBAM 模块,分别从通道和空间两个维度添加注意力机制,AP 提升了 1.26%, t 缩短 0.003 s。V3 在 V2 的基础上添加改进后的 BiFPN 替换原来的 PANet,AP 提升了 2.37%, t 缩短 0.001 s。V4 在 V3 的基础上使用 CIUO 替换原模型的 GIOU 损失函数,AP 提升了 1.25%, t 缩短 0.004 s。V4 相比于 YOLOv5 模型,AP 提升了 5.13%, t 缩短了 0.01 s。由表 1 可知,分别添加不同的模块都会给模型带来性能的提升,且这些增益可以互补,将它们结合在一起能进一步增强网络的性能。

表 1 YOLOv5 添加了各个模块后的性能对比

模型	AP/%	t/s
YOLOv5s	93.66	0.052
V1	93.91	0.048
V2	95.17	0.045
V3	97.54	0.046
V4	98.79	0.042

4) 与其他目标检测算法的对比实验

实验对比了改进后的 YOLOv5s 模型与当前主流的目标检测算法在电表示数检测中的性能。本文模型在主干网络中添加了 CBAM 模块,在特征金字塔部分使用 D-BiFPN 替代原来的 PANet,使用 CIUO 优化边界框损失函数。表 2 所示为不同的目标检测算法在电表示数检测任务中的性能,可知本文所提出的算法在检测精度上仅仅比 Faster R-CNN 低 0.58%,但是单张图像的检测速度却是 Faster R-CNN 的约 26 倍,SSD 的约 19 倍,综合性能最好,更适合于电表示数的目标检测任务。

表 2 不同检测算法比较

模型	AP/%	t/s
YOLOv5s	93.66	0.052
Faster R-CNN	94.24	1.089
SSD	87.37	0.827
CTPN	83.38	0.848
EAST	79.49	0.085
V4	98.79	0.042

3.2 电表示数识别实验结果分析

与光学字符识别相比,电表示数识别具有文本占比少、文本数量少和背景复杂的特点。和自然场景文本识别不同的是,电表示数连续的几个数字是实验所需要的正样本,其他数字、汉字和英文字符都是干扰文本,作为实验中的负样本处理。电表示数长短不一,数显屏幕背光颜色多样,都对识别有一定干扰。

1) 网络训练

电表示数识别所用数据集是根据电表数据集标注文

件中的文本的坐标信息把原始图像裁剪成本图像的,共裁剪出了 5 000 张图像。之后通过等比例缩放把图像的高度调至 32 个像素,以符合 CRNN 对于输入图像的要求。将数据集按照 8 : 1 : 1 的比例分成训练集、验证集和测试集。

示数识别网络训练 epoch 设置为 1 000,初始学习率为 0.001, Batch size 为 8,采用 Adam 优化器来更新梯度。

2) 字典

在生成识别数据的同时需要给图像打上对应的文本内容标签,使用字典检索的方式。字典中包含 0~9 和小数点,另外加入 CTC 中使用到的 blank 字符,共 12 个字符。

3) 算法改进前后性能对比

本文选用文本识别准确率和检测一张图片消耗的时间作为评判识别效果的指标,文本识别准确率的定义为式(15):

$$F = \frac{P}{T} \times 100\% \quad (15)$$

式中: F 为文本识别率, P 为被正确识别出来的样本数, T 为整体样本数。 F 越大,表明算法的识别效果越好。

对 CRNN 网络更换不同的主干网络实验结果如表 3 所示。

表 3 不同主干网络的性能对比

模型	F/%	T/s
VGG16	91.4	0.054
Resnet	93.3	0.057
Mobilenet v3	89.1	0.049
DenseNet	98.8	0.061

由表 3 中的实验结果可知,在 CRNN 中替换不同的 backbone,网络性能具有不同效果的提升,CRNN-Rsnet 的文本识别率相比于原 CRNN 率提升了 1.9%,CRNN-Densenet 的文本识别率提升了 7.4%,模型的性能能够达到实际的应用需求。将识别结果和输入融合处理后输出的部分实验结果如图 10 所示。



图 10 部分实验结果示意图

4 结 论

本文基于改进的 YOLOv5 与 CRNN 算法设计了一种电表示数识别系统,实现了智能电表的自动检测与识别功

能,本论文的改进点如下:

在网络层之间引入空间和通道维度的注意力机制,使网络可以自主关注图像中的重要特征。

由于电表示数目标区域过小,为了防止因目标过小而导致的漏检问题,提出 D-BiFPN 网络,进一步添加卷积操作获取更深层语义信息,并且对提取的深层语义特征进行融合,以提升相邻特征间的利用率和联系,增强了网络的检测能力,使得在电表示数检测任务中漏检的问题得以改善。

引入 CIOU 替换 GIOU 损失函数,有效提高了文本检测的准确率,降低了检测时间。

使用 DenseNet 作为 CRNN 文本识别模型的主干提取网络,大幅提高了检测精度。

算法在电表图像示数检测识别具有良好的鲁棒性,克服了人工读取示数时费时、费力和错误率高的缺点。本文立足于工程实际应用需求,满足变电站巡检和电表自动化检定等应用环境的需求,模型参数量小,易于部署在移动终端。示数检测的准确率达到 98.79%,示数识别的准确率达到 98.8%,达到了预期目的。

参考文献

- [1] 欧家祥,史文彬,张俊玮,等. 基于深度学习的高效电力部件识别[J]. 电力大数据, 2018, 9(21): 1-8.
- [2] ZHOU X Y, YAO C, WEN H, et al. EAST: an efficient and accurate scene text detector [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA. IEEE, 2017: 2642-2651.
- [3] LIN L, LU Z H. Automatic character recognition method of power meter based on vision [C]. 3rd International Conference on Electronic Information Technology and Computer Engineering. Xiamen, China: Institute of Electrical and Electronics Engineers Inc, 2019: 1667-1670.
- [4] ZHAN F N, LU S J. ESIR: end-to-end scene text recognition via iterative image rectification [C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA. IEEE, 2019: 2054-2063.
- [5] BELHARBI S, HÉRAULT R, CHATELAIN C, et al. Deep neural networks regularization for structured output prediction [J]. Neurocomputing, 2018, 281: 169-177.
- [6] YANG F, JIN L, LAI S, et al. Fully convolutional sequence recognition network for water meter number reading [J]. IEEE Access, 2019, 7: 11679-11687.
- [7] TIAN Z, HUANG W L, HE T, et al. Detecting text in natural image with connectionist text proposal

- network[C]. Computer Vision-ECCV 2016, 2016: 224-232.
- [8] ZHENG Z H, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.
- [9] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37(9): 1904-1916.
- [10] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 936-944.
- [11] 杨知, 欧文浩, 费香泽, 等. 基于高分辨率 SAR 影像和深度学习的输电杆塔智能识别研究[J]. 电测与仪表, 2020, 57(4): 7.
- [12] 郭敬东, 陈彬, 王仁书, 等. 基于 YOLO 的无人机电力线路杆塔巡检图像实时检测 [J]. 中国电力, 2019, 52(7).
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), Munich: Springer, 2018: 3-19.
- [14] 张焕坤, 李军毅, 张斌, 等. 基于改进型 YOLOv3 的绝缘子异物检测方法[J]. 中国电力, 2020, 53(2): 49-55.
- [15] 周仿荣, 方明等. 基于 YOLOv3 的输电线路缺陷快速检测方法[J]. 云南电力技术, 2020, 48(4): 6.
- [16] ZHANG J, JIANG F. Multi-level supervised network for person re-identification [C]. Proceedings of the 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Piscataway: IEEE, 2019: 2072-2076.
- [17] 卢来, 邓文, 吴卫祖. 基于改进深度卷积神经网络的步态识别算法[J]. 电子测量与仪器学报, 2019(2): 6.
- [18] 陈同凡, 刘云鹏, 裴少通. 基于改进 YOLOv3 的电力设备红外诊断方法[J]. 广东电力, 2021, 34(6): 9.
- [19] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2017, 39(11): 2298-2304.
- [20] 姚万金, 李金平. 基于改进的 Faster R-CNN 的电力部件识别[J]. 电力科学与工程, 2019(5): 5.
- [21] 高强, 阳武, 李倩. 基于稀疏差异深度信念网络的绝缘子故障识别算法[J]. 电测与仪表, 2016, 0(1): 19-25.
- [22] 周自强, 陈强, 马必焕. 一种改进的 YOLO 目标检测方法在电缆设备异常状态识别中的应用[J]. 电测与仪表, 2020, 57(2): 7.
- [23] 公茂法, 接怡冰, 李美蓉, 等. 基于小波包与改进的 PSO-PNN 变压器励磁涌流识别算法研究[J]. 电测与仪表, 2018(8): 7.
- [24] 周克, 杨倩文, 等. 一种改进的压板状态识别 SSD 算法[J]. 电测与仪表, 2021, 58(1): 69-76.

作者简介

黄辉, 副教授, 主要研究方向为机器视觉和电气设备故障在线监测。

肖豪(通信作者), 硕士研究生, 主要研究方向为人工智能与机器视觉技术。

E-mail: 1078543268@qq.com