



DOI: 10.12086/oe.2021.200094

## 基于神经网络的车辆识别代号识别方法

孟凡俊<sup>1,2</sup>, 尹东<sup>1,2\*</sup><sup>1</sup>中国科学技术大学信息科学技术学院, 安徽 合肥 230027;<sup>2</sup>中国科学院电磁空间信息重点实验室, 安徽 合肥 230027

**摘要:** 在车辆识别和车辆年检时, 正确识别车架上金属刻印的车辆识别代号(VIN)是非常重要的环节。针对 VIN 序列, 本文提出了一种基于神经网络的旋转 VIN 图片识别方法, 它由 VIN 检测和 VIN 识别两部分组成。首先, 在 EAST 算法基础上利用轻量级神经网络提取特征, 并结合文本分割实现快速、准确的 VIN 检测; 其次, 将 VIN 识别任务作为一个序列分类问题, 提出了一种新的识别 VIN 方法, 即通过位置相关的序列分类器, 预测出最终的车辆识别代号。为了验证本文的识别方法, 引入了一个 VIN 数据集, 其中包含用于检测的原始旋转 VIN 图像和用于识别的水平 VIN 图像。实验结果表明, 本文方法能有效地识别车架 VIN 图片, 同时达到了实时性。

**关键词:** 车辆识别代号; 神经网络; 文本分割; 机器视觉

**中图分类号:** TP391.41

**文献标志码:** A

孟凡俊, 尹东. 基于神经网络的车辆识别代号识别方法[J]. 光电工程, 2021, 48(1): 200094

Meng F J, Yin D. Vehicle identification number recognition based on neural network[J]. *Opto-Electron Eng* 2021, 48(1): 200094

## Vehicle identification number recognition based on neural network

Meng Fanjun<sup>1,2</sup>, Yin Dong<sup>1,2\*</sup><sup>1</sup>School of Information Science Technology, University of Science and Technology of China, Hefei, Anhui 230027, China;<sup>2</sup>Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei, Anhui 230027, China

**Abstract:** It is far essential to properly recognize the vehicle identification number (VIN) engraved on the car frame for vehicle surveillance and identification. In this paper, we propose an algorithm for recognizing rotational VIN images based on neural network which incorporates two components: VIN detection and VIN recognition. Firstly, with lightweight neural network and text segmentation based on EAST, we attain efficient and excellent VIN detection performance. Secondly, the VIN recognition is regarded as a sequence classification problem. By means of connecting sequential classifiers, we predict VIN characters directly and precisely. For validating our algorithm, we collect a VIN dataset, which contains raw rotational VIN images and horizontal VIN images. Experimental results show that the algorithm we proposed achieves good performance on VIN detection and VIN recognition in real time.

**Keywords:** vehicle identification number; neural network; text segmentation; machine vision

收稿日期: 2020-03-20; 收到修改稿日期: 2020-06-28

基金项目: 安徽省 2018 年度重点研究与开发计划项目(1804a09020049)

作者简介: 孟凡俊(1994-), 男, 硕士研究生, 主要从事图像处理的研究。E-mail: fanjunm@mail.ustc.edu.cn

通信作者: 尹东(1965-), 男, 副教授, 主要从事图像处理的研究。E-mail: yindong@ustc.edu.cn

版权所有©2021 中国科学院光电技术研究所



当网络越深入,网络越难训练,越难收敛,ResNet<sup>[4]</sup>采用残差块连接不同层次的神经网络进行深度学习,进一步提高了神经网络性能。

## 2.2 文字检测

传统的文本检测方法主要是两种:基于连通域分析的方法和基于滑动检测窗的方法。这些方法首先获得文本候选区域,然后采用手工设计的特征对候选区域分析,提取文本定位。其中,基于连通域的方法,例如 MSER<sup>[5-7]</sup>、SWT<sup>[8]</sup>等采用自底向上的方法,从边缘以及像素的角度出发,最后连接在一起。在论文<sup>[9]</sup>中对图像使用直方图均衡化和二值化处理结合,连通域分析找出字符区域。而基于滑动检测窗<sup>[10-11]</sup>的方法自顶向下利用滑动检测窗的方式对整幅图像进行扫描,获得候选区域。

由于手工设计特征的分类能力不足,随着深度学习的兴起,近些年出现了很多基于神经网络的文本检测方法,这些方法通过神经网络学习和深度特征,从而快速而有效地检测图像中的文本。基于深度学习的文本检测方法主要包括基于区域建议的方法和基于分割的方法。基于区域建议方法一般遵循目标检测的框架,例如 CTPN<sup>[12]</sup>的方法在 Faster RCNN<sup>[13]</sup>基础上采用了竖直的小的锚点框去预测连续的文本区域,最后连接在一起;TextBoxes<sup>[14]</sup>采用修改 SSD<sup>[15]</sup>目标检测框架中锚点框的大小和长宽比,同时修改了卷积核的大小来检测水平文本,取得了不错的效果;TextBoxes++<sup>[16]</sup>进一步扩展了文本框的文本检测性能,实现了对旋转文本的检测;SegLink<sup>[17]</sup>基于 SSD 检测框架,同时预测文本片段和文本连接关系,并在网络中加入角度信息,对任意角度的文本进行检测;RRPN<sup>[18]</sup>在 Faster RCNN 基础上添加了旋转角度的锚点框,以解决检测任意角度文本的问题。同样,近年来也出现了许多基于图像分割的文本检测方法,图像分割试图在像素级基础上

对每一个像素分配一个标签以实现分割不同区域。EAST<sup>[19]</sup>将 FCN<sup>[20]</sup>网络方法与像素分割相结合,直接预测文本的边界;PixelLink<sup>[21]</sup>采用了一种新的文本检测方法,它可以预测像素是否属于文本,以及文本像素之间的连接,将同一实例中的像素连接起来,分离不同的文本实例,然后直接从分割结果中提取文本框。

## 2.3 文字识别

以前的很多文字识别方法首先检测文本行中的每个字符,然后将每个字符单独分割,最后通过自下而上的拼装方法将字符组合成一个单词或句子。近些年来,随着深度学习的兴起,逐渐将文本识别看作一个序列化识别的问题。CRNN<sup>[22]</sup>将文本识别作为一个序列化识别问题,首先利用深度卷积神经网络提取文本特征,然后学习文本的空间上下文信息,最后利用 CTC 层<sup>[23]</sup>对文本序列进行解码;论文<sup>[24]</sup>提出了一种基于注意力机制的序列到序列框架来识别文本。通过这种方式,神经网络能够从训练数据中学习隐藏在字串中的字符级语言模型。

## 3 本文方法

本文提出了一种基于神经网络的旋转 VIN 图片识别方法。整体框架如图 2 所示。

### 3.1 VIN 图片文字检测

受到之前发布的目标检测和文本检测论文方法的启发,本文在 EAST 的算法上采用 MobileNet<sup>[25]</sup>作为基础网络,这是一种轻量级的网络,使得在性能损失很少的情况下推理速度加快,模型参数更少,能满足识别任务的实时性要求。为了检测不同大小和比例的 VIN 图片,采用与特征金字塔网络(FPN)<sup>[26]</sup>相似的方法,将低级特征图和高级语义特征图连接起来。FPN 是一种能够提取多尺度特征图进行融合的特征提取网络,从而提高了检测性能。同时,本文通过将由下而

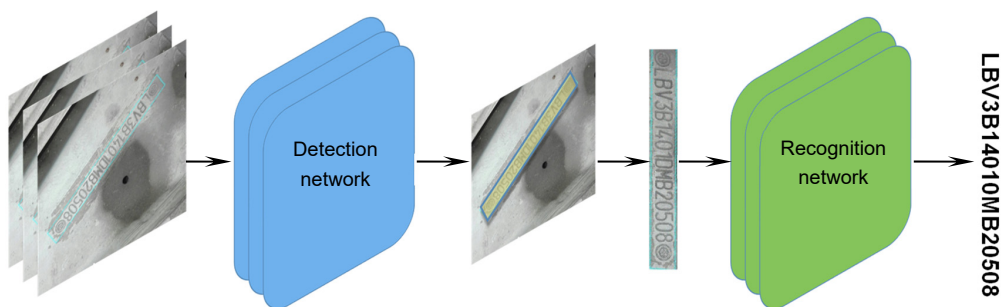


图 2 基于神经网络的 VIN 识别框架

Fig. 2 The overall architecture of our proposed VIN recognition algorithm

上的网络横向连接来学习一个金字塔型强语义特征，如图 3 所示。

为了实现更精细化的文字区域预测，在卷积神经网络抽取输入图片特征之后，全卷积网络 FCN 在最后一层卷积层上实现像素级的预测。全卷积网络是一种用于图像语义分割的网络，这种网络实现像素级的目标分割，同时不受图片输入大小的限制。

文字检测可以作为一个二元分类问题。在最终的预测结果中，如图 4 所示，输出层的一个通道  $C_1$  输出预测的像素在文本区域内的概率，同时设置一个阈值来滤除可能位于背景区域内的概率较低的像素。对于预测在文本区域内的每个像素，有 4 个通道  $C_2, C_3, C_4, C_5$  输出预测的该像素到文本段四条边的距离。最后，还有一个通道  $C_6$  输出预测的倾斜文本区域内的方向。

VIN 检测模型损失函数包括 VIN 文本分类损失和 VIN 边界计算损失。由于 VIN 文本与背景像素分布不平衡，背景像素远远大于文本像素，故采用 dice<sup>[27]</sup>系数来计算 VIN 文本分类损失，dice 系数公式为

$$L_{cls} = 1 - \frac{2|C_1 \cap Y|}{|C_1| + |Y|}, \quad (1)$$

其中：通道  $C_1$  为预测的像素在文本区域内的概率， $Y$  为训练时图片中 VIN 文本区域的真实标签。

VIN 检测模型边界损失函数包括倾斜角度损失和边界框回归损失，如式(2)~式(4)所示：

$$L_{angle} = 1 - \cos(\theta - \theta^*), \quad (2)$$

$$L_{bbox} = \frac{B \cap B^*}{B \cup B^*}, \quad (3)$$

$$L_{detection} = L_{cls} + \lambda(L_{bbox} + L_{angle}), \quad (4)$$

其中： $L_{angle}$  是倾斜角度损失函数， $\theta$  是预测的倾斜角度， $\theta^*$  是真实的倾斜角度； $L_{bbox}$  是边界框损失函数， $B$  是预测的边界， $B^*$  是真实的边界。最终，整体的损失函数用式(4)表示。

### 3.2 VIN 图片文字识别

考虑到文本在图像中的任意方向，首先检测图像中的文本，然后将文本旋转到水平位置后将文本区域裁剪出来，进行下一步的识别。

受到 CRNN 的启发，本文采用了一种新的 VIN 识别方法，即把 VIN 的识别当作序列相关的分类，同时直接忽略不想识别的其他边缘字符，例如有的 VIN 图片中含“\*”等不相关字符，如图 5(a)所示。由于

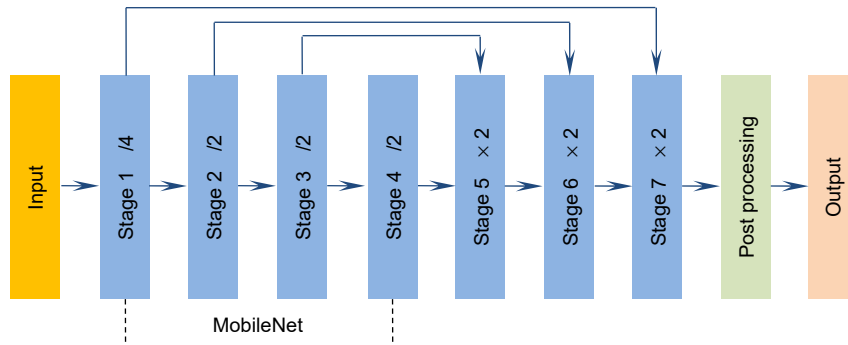


图 3 VIN 检测网络基本框架

Fig. 3 Illustration of the VIN detection architecture

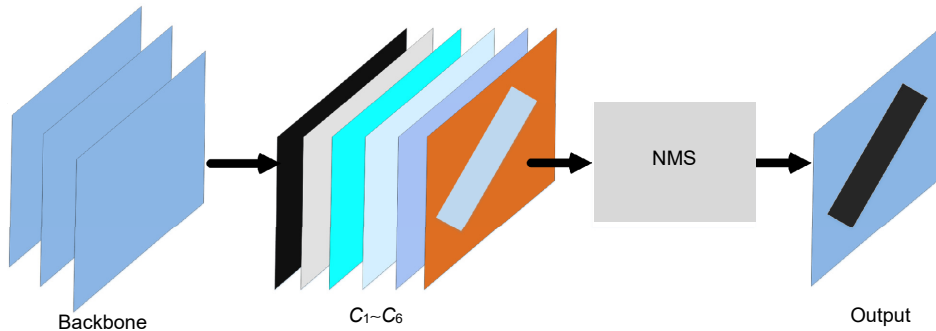


图 4 VIN 检测网络中的后处理以及输出

Fig. 4 Post-processing and output of VIN detection network

VIN 的固定长度为 17 位字符，在输出端设置 17 个分类器，输入网络的图像的大小固定为 400×400，提取特征后得到输入大小的 1/16 的特征图。图 5(b)展示了用于 VIN 特征提取的主干网络结构 VGGNet，3×3 卷积核用于提取图片特征，Maxpool 层是一个特征池化层，主要目的是 2 倍下采样，本文中网络在下采样时采用向下取整。接下来，计算 VIN 识别网络的损失函数如式(5)、式(6)所示：

$$L_{\text{rec}} = \sum_{k=1}^M L_{\text{cha}}(y_k, y_k^*) \quad (5)$$

$$L_{\text{cha}}(y_k, y_k^*) = -\sum_{i=1}^N y_k^i * \log \hat{y}_k^i \quad (6)$$

其中：M 是 VIN 的总长度，设为 17。N 是字符的所有类别，N 为 33。对于第 k 个字符， $L_{\text{cha}}(y_k, y_k^*)$  计算预测值  $y_k$  和真实标签  $y_k^*$  的交叉熵损失， $L_{\text{rec}}$  是所有的分类器的损失和。

### 3.3 数据增强

在训练 VIN 检测网络时，对训练图片以任意方向旋转以提升检测多角度图片的效果，如图 6 所示，同时旋转原图片和四边形文本框标签。

为了达到无损地旋转图片，采用仿射变化进行操作。假设原图片大小为  $H \times W$ ，任意旋转角为  $\theta(\theta \in [0, 2\pi))$ ，同时训练图片中的文本框标签为  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ ，这是一个四边形四个顶点的坐标值，仿射变化式如下：

$$H_r = W \times \sin\theta + H \times \cos\theta \quad (7)$$

$$W_r = H \times \sin\theta + W \times \cos\theta \quad (8)$$

$$R(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (9)$$

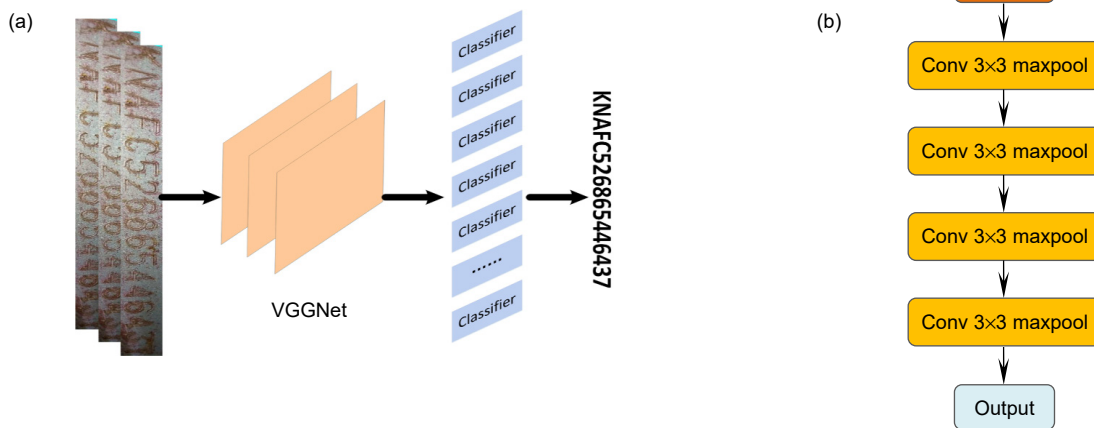


图 5 VIN 识别算法网络。(a) 网络整体框架；(b) VGGNet 内部结构  
Fig. 5 Illustration of VIN recognition network. (a) Overall framework; (b) VGGNet network

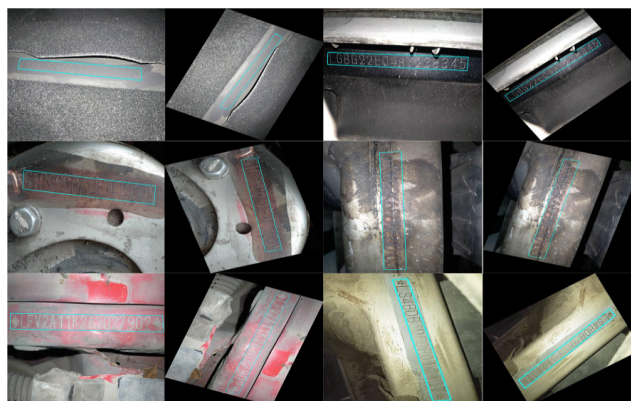


图 6 任意方向旋转图片和文本框标签  
Fig. 6 Rotating VIN images and labels at any angle to augment dataset

$$T(t_x, t_y) = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{bmatrix}, \quad (10)$$

$$M(\theta) = T\left(\frac{H}{2}, \frac{W}{2}\right) R(\theta) T\left(-\frac{H}{2}, -\frac{W}{2}\right), \quad (11)$$

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = M(\theta) \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (12)$$

其中： $H_r$  和  $W_r$  是旋转后的新的图片大小， $R(\theta)$  是旋转矩阵， $T(t_x, t_y)$  是变换矩阵，最终的仿射变换用式(12)表示。

在识别阶段，检测定位旋转到水平之后，有些文字可能方向是反的，也就是旋转了  $180^\circ$ ，为了识别旋转了  $180^\circ$  的图片，在训练识别网络的时候以 50% 的概率随机旋转图片  $180^\circ$ 。

## 4 实验结果与数据分析

本文提出了针对复杂环境下旋转 VIN 图片的识别算法，在基于 Ubuntu 16.04 系统 GPU(Titan Xp 显卡)服务器的环境上开展了训练和测试实验。实验所需的编程语言为 python3.6 版本，深度学习环境为 pytorch0.4 版本，CPU 型号为 Intel Silver 4110 2.1 GHz。

### 4.1 数据集

为了验证 VIN 识别算法，手动收集了一个数据集——车架号识别数据集，如图 7 所示。该数据集第一部分包含 2000 张旋转 VIN 图片，每张图片都有手动标注的四边形框顶点坐标。第二部分包含 90000 张水平 VIN 图片，每张图片都有对应的真实字符标签。在第一部分中，VIN 图片具有复杂的背景、不同的拍摄视角、具有任意方向和模糊的文本，文本区域的标签用四边形的四个顶点坐标表示  $(x_1, y_1, x_2, y_2, x_3, y_3, x_4, y_4)$ 。在第二部分中，水平 VIN 图像可以专门用于 VIN 识别训练。这些水平的 VIN 图片都具有复杂的背景，不同的字体以及模糊的文字，所有的图片都有 17 个 VIN 字符的真实标签。

### 4.2 实验结果与数据分析

在 VIN 检测阶段，用旋转 VIN 数据集进行实验，使用 1000 张图片用于训练，1000 张用于测试。在模型训练中，使用 Adam 作为优化器，并将训练图片裁剪为  $(512 \times 512)$  的大小。此外，在训练时随机以任意角度无损旋转图片。实验结果如表 1 所示，本文所提出的检测方法在速度和精度上均优于其它方法，精度可达 98%。



图 7 VIN 训练数据集

Fig. 7 Samples of VIN dataset

表 1 在 VIN 数据集上不同方法的检测效果

Table 1 Detection performance comparison of different algorithms on VIN dataset

Methods	FPS(frame/s)	Precision/%	Recall/%
PSENet <sup>[28]</sup>	1	86.4	92.2
EAST	20	94.3	95.7
FOTS <sup>[29]</sup>	24	89.9	93.6
OURS	33	98.5	98.5

在 VIN 识别阶段,以 80000 张图片用于训练、10000 张用于测试。识别结果如表 2 所示,本文提出的 VIN 识别算法在两个精度指标上比其他算法具有更好的识别效果,达到了 122 f/s 的识别速度和 93% 相似度匹配的识别精度。这里测试图片大小统一为 320×32,‘F’代表全字符匹配的准确率,‘A’代表字符相似匹配的准确率。

此外,值得注意的是,输入图片的不同大小在 VIN

识别中的实验效果是不同的。如表 3 所示,图像越大,实验性能越好。同时,针对数据增强,也做了对比实验。实验结果表明,对于图片中文字方向反的情况,数据增强可以有效解决这一问题。表中‘aug’表示我们在训练识别网络时添加了图片 180° 旋转。

作为补充,图 8 展示了本文方法在实际复杂环境下的 VIN 图片的识别效果。图 9 展示了在测试中检测失败以及识别率低于 0.7 的图片。经过多张图片的分

表 2 在 VIN 数据集上不同方法的识别效果

Table 2 Recognition performance comparison of two algorithms on VIN dataset

Methods	FPS(frame/s)	F/%	A/%
Tesseract(320×32) <sup>[30]</sup>	6	4.47	11.96
CRNN(320×32)	40	43.61	83.44
ASTER(320×32) <sup>[31]</sup>	33	44.25	85.19
OURS(320×32)	122	61.47	93.28

表 3 在不同尺寸输入图片大小上的识别效果

Table 3 Recognition performance comparison of different input size

Input size	FPS(frame/s)	F/%	A/%
200×20+aug	134	47.53	89.66
200×20	134	41.21	85.73
320×32+aug	122	61.47	93.28
320×32	122	55.15	91.90
400×40+aug	94	66.29	95.12
400×40	94	58.70	91.42
500×50+aug	88	64.73	95.71
500×50	88	58.11	92.37



图 8 实际复杂环境下 VIN 图片识别效果

Fig. 8 Samples of detection and recognition results of our algorithm on complicated VIN images



图9 实际复杂环境下检测和识别效果不好的图片

Fig. 9 Samples of failed detection and recognition on complicated VIN images

析,发现图片中的长条格状物体容易被错误识别为VIN文字序列,实际上只是类似于文字。而且在文字和背景没有明显区分的时候很容易造成检测不到文字的存在。在识别方面,显然模糊刻印的文字以及不明显的字体是造成识别率低的原因。

## 5 结论

本文提出了在复杂环境下识别旋转型VIN图片的有效方法,并且创建了一个新的VIN数据集来评估我们的方法。结合轻量级卷积神经网络和输出层的逐像素分割实现了VIN检测。同时,提出了用于VIN识别的序列分类器。实验结果表明,本文所提出的方法能有效地检测和识别图片中刻印的VIN字符。目前,本文方法已经应用在中国的城市车辆年检系统中。

## 参考文献

- [1] Smith R. An overview of the Tesseract OCR engine[C]// *Proceedings of the 9th International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2007: 629–633.
- [2] Mori S, Suen C Y, Yamamoto K. Historical review of OCR research and development[J]. *Proc IEEE*, 1992, **80**(7): 1029–1058.
- [3] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[Z]. arXiv:1409.1556, 2014.
- [4] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 770–778.
- [5] Tang Y B, Bu W, Wu X Q. Natural scene text detection based on multi-level MSER[J]. *J Zhejiang Univ (Eng Sci)*, 2016, **50**(6): 1134–1140.
- [6] 唐有宝, 卜巍, 郭向前. 多层次MSER自然场景文本检测[J]. *浙江大学学报(工学版)*, 2016, **50**(6): 1134–1140.
- [7] Jiang H Y, Zhu L P, Ou Z P. Text recognition of natural scene image based on MSER and Tesseract[J]. *Comput Knowl Technol*, 2017, **13**(33): 213–216.
- [8] 蒋弘毅, 朱丽平, 欧樟鹏. 基于MSER和Tesseract的自然场景图像文字识别[J]. *电脑知识与技术*, 2017, **13**(33): 213–216.
- [9] Zhang K Y, Shao K Y, Lu D. MSER fast skewed scene-text location algorithm[J]. *J Harbin Univ Sci Technol*, 2019, **24**(2): 81–88.
- [10] 张开玉, 邵康一, 卢迪. MSER快速自然场景倾斜文本定位算法[J]. *哈尔滨理工大学学报*, 2019, **24**(2): 81–88.
- [11] Zhang G H, Huang K, Zhang B, et al. A natural scene text extraction method based on the maximum stable extremal region and stroke width transform[J]. *J Xi'an Jiaotong Univ*, 2017, **51**(1): 135–140.
- [12] 张国和, 黄凯, 张斌, 等. 最大稳定极值区域与笔画宽度变换的自然场景文本提取方法[J]. *西安交通大学学报*, 2017, **51**(1): 135–140.
- [13] Nan Y, Bai R L, Li X. Application of convolutional neural network in printed code characters recognition[J]. *Opto-Electron Eng*, 2015, **42**(4): 38–43.
- [14] 南阳, 白瑞林, 李新. 卷积神经网络在喷码字符识别中的应用[J]. *光电工程*, 2015, **42**(4): 38–43.
- [15] Wang K, Belongie S. Word spotting in the wild[C]// *Proceedings of the 11th European Conference on Computer Vision*, 2010, **6311**: 591–604.
- [16] Wang K, Babenko B, Belongie S. End-to-end scene text recognition[C]// *Proceedings of 2011 International Conference on Computer Vision*, 2011: 1457–1464.
- [17] Tian Z, Huang W L, He T, et al. Detecting text in natural image with connectionist text proposal network[C]// *Proceedings of the 14th European Conference on Computer Vision*, 2016, **9912**: 56–72.



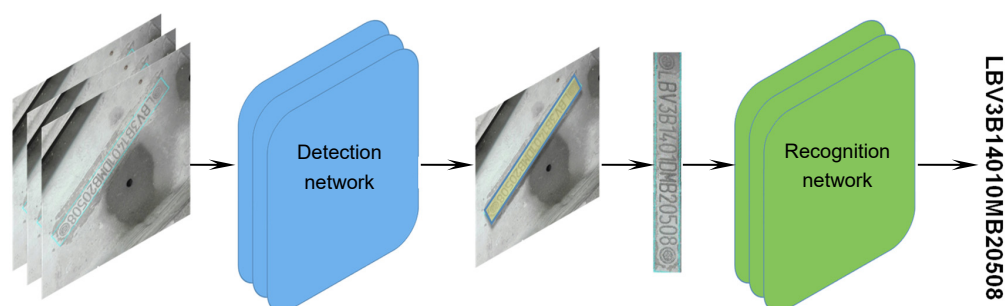
- [13] Ren S Q, He K M, Girshick R, *et al.* Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.
- [14] Liao M H, Shi B G, Bai X, *et al.* TextBoxes: a fast text detector with a single deep neural network[Z]. arXiv:1611.06779, 2016.
- [15] Liu W, Anguelov D, Erhan D, *et al.* SSD: single shot multibox detector[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016: 21–37.
- [16] Tian Z, Huang W L, He T, *et al.* Detecting text in natural image with connectionist text proposal network[C]//*Proceedings of the 14th European Conference on Computer Vision*, 2016, **9912**: 56–72.
- [17] Liao M H, Shi B G, Bai X. TextBoxes++: a single-shot oriented scene text detector[J]. *IEEE Trans Image Process*, 2018, **27**(8): 3676–3690.
- [18] Ma J Q, Shao W Y, Ye H, *et al.* Arbitrary-oriented scene text detection via rotation proposals[J]. *IEEE Trans Multimed*, 2018, **20**(11): 3111–3122.
- [19] Zhou X Y, Yao C, Wen H, *et al.* EAST: an efficient and accurate scene text detector[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 2642–2651.
- [20] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//*Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015: 3431–3440.
- [21] Deng D, Liu H, Li X, *et al.* PixelLink: detecting scene text via instance segmentation[Z]. arXiv:1801.01315, 2018.
- [22] Shi B G, 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 Trans Pattern Anal Mach Intell*, 2017, **39**(11): 2298–2304.
- [23] Graves A, Fernández S, Gomez F, *et al.* Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]//*Proceedings of the 23rd International Conference on Machine Learning*, 2006: 369–376.
- [24] Lee C Y, Osindero S. Recursive recurrent nets with attention modeling for OCR in the wild[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016: 2231–2239.
- [25] Sandler M, Howard A, Zhu M L, *et al.* MobileNetV2: inverted residuals and linear bottlenecks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 4510–4520.
- [26] Lin T Y, Dollár P, Girshick R, *et al.* Feature pyramid networks for object detection[C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017: 936–944.
- [27] Milletari F, Navab N, Ahmadi S A. V-Net: fully convolutional neural networks for volumetric medical image segmentation[C]//*Proceedings of the 4th International Conference on 3D Vision (3DV)*, 2016: 565–571.
- [28] Li X, Wang W H, Hou W B, *et al.* Shape robust text detection with progressive scale expansion network[Z]. arXiv:1806.02559, 2018.
- [29] Liu X B, Liang D, Yan S, *et al.* FOTS: fast oriented text spotting with a unified network[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 5676–5685.
- [30] Thakare S, Kamble A, Thengne V, *et al.* Document Segmentation and Language Translation Using Tesseract-OCR[C]//*2018 IEEE 13th International Conference on Industrial and Information Systems (ICIIIS)*. IEEE, 2018.
- [31] Shi B G, Yang M K, Wang X G, *et al.* ASTER: an attentional scene text recognizer with flexible rectification[J]. *IEEE Trans Pattern Anal Mach Intell*, 2019, **41**(9): 2035–2048.

# Vehicle identification number recognition based on neural network

Meng Fanjun<sup>1,2</sup>, Yin Dong<sup>1,2\*</sup>

<sup>1</sup>School of Information Science Technology, University of Science and Technology of China, Hefei, Anhui 230027, China;

<sup>2</sup>Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Hefei, Anhui 230027, China



The overall architecture of our proposed VIN recognition algorithm

**Overview:** It is far essential to properly recognize the vehicle identification number (VIN) engraved on the car frame for car surveillance and vehicle identification. Vehicle identification number is unique globally, which is assigned by car manufacturers to a car for identifying it. The vehicle identification number is usually engraved on the metallic car frame which is uneasy to tamper with, so it is of great significance for vehicle annual surveillance and vehicle identification. Although many important achievements have been made in text recognition, especially the wide application of OCR in document recognition in images, the complex background, arbitrary angle and fuzzy font of the engraved text in the images have made it difficult to identify the vehicle identification number automatically. In vehicle identification and annual car inspection, a large number of VIN pictures need to be manually reviewed every day, which is very inefficient. With the application of deep learning, we can make use of deep learning to accelerate this process, improve the efficiency of auditing greatly, and realize automated auditing. We introduce an algorithm for recognizing vehicle identification number in images based on neural network, which incorporates two components: VIN detection and VIN recognition. Firstly, in the VIN detection part, the lightweight Network is used as feature extraction network in order to accelerate the inference speed and reduce the network cost. Combined with FCN and FPN, the network is able to adapt to any size of input images and focus on the distribution difference between foreground text pixels and background pixels. In order to improve the performance on rotational VIN, the images are rotated at any angle lossless in the training stage to augment datasets. Secondly, in the VIN recognition stage, we take VIN recognition task as a sequence classification problem, using VGGNet as the feature extraction network, and the final vehicle identification number sequence is predicted through the position-related sequential classifier without character segmentation to simplify the recognition processing. Also, the text direction in images may be reversed in dataset, and in order to solve the situation, picture is rotated at 180 degrees randomly in network training. Finally, we introduce a VIN dataset, which contains raw rotational VIN images and horizontal VIN images for validating our algorithm, and all of our experiments are conducted on the dataset. Experimental results show that the algorithm we proposed can detect and recognize the VIN text in images efficiently in real time.

Meng F J, Yin D. Vehicle identification number recognition based on neural network[J]. *Opto-Electron Eng*, 2021, **48**(1): 200094; DOI: 10.12086/oe.2021.200094

Foundation item: Key Research and Development Projects of Anhui Province (804a09020049)

\* E-mail: yindong@ustc.edu.cn