基于卷积神经网络的铁轨路牌识别方法

孟 琭[1] 孙霄宇[1] 赵 滨[2] 李 楠[3]

摘 要 轨道交通在我国综合交通运输体系中占有重要的地位, 随着人工智能的发展, 智能感知轨道交通周围环境的信息也变得越来越引人注目. 本文结合深度学习与图像处理的方法, 设计并实现了一种基于卷积神经网络的高铁轨道周边路牌数字识别的智能系统, 该系统通过在高铁驾驶室内安装摄像头的方式采集运行前方的视频, 并通过目标识别、语义分割等深度学习算法自动定位并识别路牌内的数字, 从而解决了之前人工处理的繁琐和低效率.本算法整体系统由三个子模块构成, 分别为目标检测模块、语义分割模块以及数字识别模块, 其中目标检测模块基于 SSD (Single shot MultiBox dector) 模型, 并对其进行了改进, 使其更适用于轨道交通中的小目标识别; 语义分割模块使用了全卷积的方式, 对目标检测的结果进一步处理, 准确得到路牌中的数字区域; 数字识别模块的设计参考了著名的识别 MNIST 数据集的手写体识别系统, 并针对路牌中数字的特点做了相应的改进, 实现了对每个数字的准确识别. 实验结果表明, 本系统可适应白天、夜间情况下轨道交通的路况, 识别的综合准确率为 80.45%, 其中, 白天的平均识别准确率为 87.98%, 夜间的平均识别准确率为 72.92%.

关键词 智能轨道交通, 高铁路牌识别, 深度学习, 图像处理, 目标检测

引用格式 孟琭, 孙霄宇, 赵滨, 李楠. 基于卷积神经网络的铁轨路牌识别方法. 自动化学报, 2020, **46**(3): 518−530

DOI 10.16383/j.aas.c190182

An Identification Method of High-speed Railway Sign Based on Convolutional Neural Network

MENG Lu[1] SUN Xiao-Yu[1] ZHAO Bin[2] LI Nan[3]

Abstract Rail transit plays an important role in China's comprehensive transportation system.Intelligent perception of environmental information around rail traffic is also becoming more and more attractive. Combining the methods of deep learning and image processing, the paper designs and implements an intelligent system that is based on convolutional neural network for identification of rail digital signs around high-speed rail. The system not only collects videos by installing a camera in the high-speed rail cab but also automatically locates and identifies the numbers in the railway sign by the depth learning algorithm such as object detection and semantic segmentation, which can solve the cumbersome and inefficient manual processing. The total system of the algorithm consists of three sub-modules: the object detection module is based on the single shot MultiBox dector (SSD) model and improves it to be more suitable to detect the small target in the rail transit; the semantic segmentation module uses the full convolution method to further process the result of the object detection module and then get accurate digital region in the rail sign; the design of the digital identification module referred to the famous handwriting recognition system that recognizes the MNIST dataset. Besides,it improved the characteristics of the numbers in the railway signs and achieved the accurate identification of each number. The experimental results show that the system can adapt to the conditions of various rail transits, including: day and night. The comprehensive accuracy of recognition is 80.45%. Furthermore,the average accuracy of the daytime is 87.98%, and the average accuracy of the night is 72.92%.

Key words Intelligent rail transit, high-speed railway sign recognition, deep learning, image processing, object detection

Citation Meng Lu, Sun Xiao-Yu, Zhao Bin, Li Nan. An identification method of high-speed railway sign based on convolutional neural network. *Acta Automatica Sinica*, 2020, **46**(3): 518−530

收稿日期 2019-03-19 录用日期 2019-08-08
Manuscript received March 19, 2019; accepted August 8, 2019
本文责任编委 阳春华
Recommended by Associate Editor YANG Chun-Hua
1. 东北大学信息科学与工程学院 沈阳 110819 2. 友和利德科技有限公司 天津 300452 3. 沈阳产品质量监督检验院 沈阳 110022
1. College of Information Science and Engineering, Northeastern University, Shenyang 110819 2. UUValue Technology Co., Ltd, Tianjin 300452 3. Shenyang Product Quality Supervision and Inspection Institute, Shenyang 110022

近些年来, 中国的轨道交通行业发展十分迅速, 城市内、城市间轨道交通建设速度逐年提高, 运营里程和路网密度大幅提高[1]. 轨道交通行业的快速发展对效率的要求越来越高, 因此人工智能在这一领域的重要性也不断体现出来[2].

本文中高铁路牌的智能识别, 就来自于某省铁路局的实际需求. 高速铁路两旁每隔几米远就会有

架设的高压输电线路保障列车有足够的动力, 而对每个输电架进行编号可以方便铁路保障人员对输电线路进行检修, 保证列车每天的正常运行. 线路维护人员会在列车驾驶室内放置摄像机拍摄铁路两旁的路牌 (如图 1 所示), 然后再通过人工的方式对每个路牌编号进行手工记录, 其工作效率十分低下. 而且由于两个输电架之间的距离比较近, 通常只有几米, 而列车的速度最低可达 120 km/h. 这使得每个路牌出现在视频中的时间通常不足 1 s, 大大增加了人工识别的困难. 在这样的背景下, 通过利用人工智能进行自动的路牌定位与识别, 输出路牌上的数字以及在视频中出现的时间, 通过输出可以知道某一个输电架的具体位置, 方便人员的维修与保障, 显著的提高工作效率, 降低人工成本.

目前, 卷积神经网络在图像检测与识别领域应用广泛. Faster R-CNN, YOLO, Mask R-CNN 等目标检测、分割算法的提出使得识别的精度与速度不断提高[3−8]. 应用卷积神经网络设计出的手写体数字识别网络的准确性也超过了传统机器学习算法[9−11]. 然而以上算法并不是面向智能轨道交通来设计的, 而本系统所要实现的目标有着轨道交通所特有的性质, 包括: 1) 路牌的尺寸相对于整张图像来说, 比例非常小, 一般小于 1 %, 这使得一般的目标检测算法识别准确率较低; 2) 语义分割需要在一个尺寸为 20×50 左右的矩形区域内, 实现精准的数字区域分割, 而语义分割算法例如 Deeplab, Mask R-CNN 算法都是在 256 像素 × 256 像素或 128 像素 × 128 像素的这样量级的矩形区域内实现比较粗线条的分割[12]; 3) MNIST 是面向手写体设计的, 而路牌中是印刷体, 因此需要重新制作数据集并且在训练时候使用小型数字识别网络来保证检测速度. 因此, 本文在实际需求的驱动下, 改进了传统的 SSD(Single shot MultiBox dector) 模型[13], 使得新模型在本文数据集中平均精度均值 (Mean average precision, mAP) 达到 80 % 以上, 检测速度 0.07 s/幅, 设计了新的语义分割模型, 同时兼顾了分割准确性以及速度和数字识别模型, 从而使得本系统适用于智能轨道交通的路牌识别.

1　方法

1.1　算法总体结构

为了实现视频中路牌数字的识别, 可选方案有三种:

1) 使用一个卷积神经网络, 通过对原图像进行检测, 将路牌中的三个数字当作一个目标进行检测, 但是这样需要 000 ∼ 999 共一千个类别, 而且需要不同尺度的路牌图片, 数据集图片数目庞大, 且不利于标注.

2) 使用一个卷积神经网络, 将路牌中每一张图片中的每个数字作为一个目标, 这样一张图片中就会识别出三个目标, 但是由于全图尺寸为 512 像素 × 512 像素, 而每个数字的尺寸在原图中只占有非常小的一块区域, 大致范围在 14 像素 × 14 像素 ∼ 24 像素 × 24 像素, 只占原图大小的0.074 % ∼ 0.022 %, 因此通过神经网络的多尺度计算后很可能特征高度语义化, 很难得到准确的空间位置信息, 导致预测位置不准确.

3) 采用三个卷积神经网络网络串联的模块化结构, 首先为目标检测模块, 此模块输入为视频中

图 1　高铁运行过程中, 驾驶室内摄像头所拍摄的视频图像路牌内包含数字, 其范围为 000 ∼ 999

Fig. 1　The image captured by the camera in the cab, the number range is 000 ∼ 999

每一帧图片, 在整张图片中定位路牌的位置; 然后, 将路牌区域送入之后语义分割模块中, 进行精确分割, 从而得到路牌中数字准确区域, 并对其进行二值化; 最后, 将二值化的数字图像送入数字识别模块, 对其进行分类, 从而实现对每个数字的十个分类, 即: 0 ~ 9.

本文采取了第 3 种模块化结构, 避免了目标特征的高度语义化, 同时减少了训练图片的数目, 使算法可以快速的训练与调优. 此外, 根据列出行驶环境的不同, 本文对白天、夜间两种条件下的路牌识别, 分别做了不同的处理, 因为通过对视频文件分析发现, 夜晚情况下由于路牌被列车灯光照射所以对比度远高于白天, 可以使用传统的阈值分割和形态学的图像处理方法来提取数字区域, 无需使用语义分割模块, 系统整体结构图如图 2 所示.

1.2 路牌检测模块

路牌检测模块检测流程如图 3 所示.

模块输入为视频中每帧图像, 通过计算平均像素值判断后续使用白天检测模型还是夜间检测模型, 检测算法是对传统的 SSD 算法进行改进, 之后将检测到的路牌区域裁剪并送入下一模块. SSD 是一种直接预测目标类别和位置的检测算法, 相比于 R-CNN 系列算法, 它的最大特点是目标检测速度快. 但是由于 SSD 算法是基于 VGG16 网络结构进行特征提取, 对高铁路牌这类小目标的检测能力不足[14]. 根据对算法的网络结构分析, 浅层特征感受野较小, 容易包含小目标特征, 但是不足以描述目标类别, 而深层特征虽然可以区分目标类别, 但是由于高度语义化, 损失了位置信息, 对小目标的位置描述不足, 因此对原始网络结构进行改进, 将浅层特征与深层特征融合, 将得到的新特征图作为算法的特征模板, 用于后续层进行运算. 特征融合的 SSD 算法的主要思想是通过对前端网络提取的深度特征图模板进行多次卷积, 在其中选取若干层的卷积结果作为预测输出, 找到优的目标边界框与类别概率. 算法具有以下几个特征:

1) 多尺度特征图. 在轨道交通视频中目标往往是由远及近, 具有不同尺度, 特征融合的 SSD 算法通过使用卷积层来进行目标检测, 这些层的尺度是不断减小的, 可以使算法在不同尺度大小上进行预测.

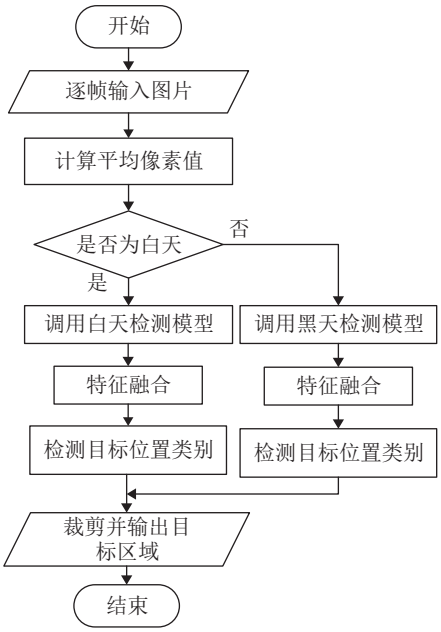


图 3 路牌检测模块算法流程图

Fig. 3 Algorithm flow chart of road sign detection module

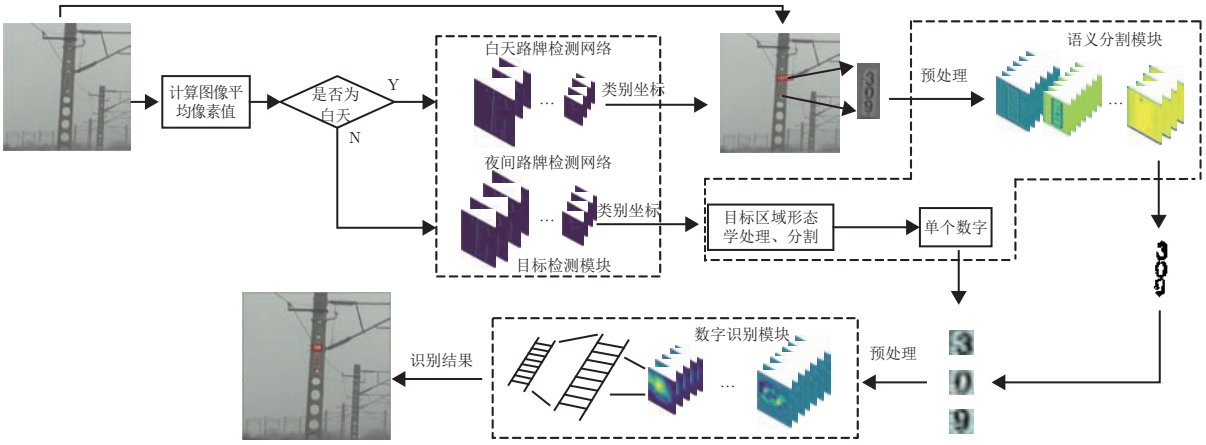


图 2 系统整体结构图

Fig. 2 Schematic of the whole system

2) 使用卷积结果进行目标检测. 算法的主网络结构是 VGG16, 用于提取图像深度特征, 之后通过将两个全连接层改成卷积层并且增加了 4 个卷积层用于预测目标的边界框与置信度 (Confidence). 算法中采用额外添加多个卷积层的目的是多重采样, 产生尽量多的边框, 这样可以大大增加边框内包含目标的概率, 提高算法预测的准确率.

3) 使用 default box 与不同长宽比. 算法采用了 anchor box 的思想[15], 将卷积层输出的特征图上每个像素映射在原图中的范围称为 default box, 算法, 即利用这些 default box 预测坐标值与置信度. 另外, 算法引入了不同的长宽比 $\alpha_r \in (1, 2, 3, \frac{1}{2}, \frac{1}{3})$, 通过将长宽比与预先设置的基础 default box 尺度比 s_k 代入如下公式

$$w_k^a = s_k\sqrt{a_r},\ h_k^a = \frac{s_k}{\sqrt{a_r}} \tag{1}$$

其中, w_k^a, h_k^a 为尺度 k 下长宽比为 a 时 default box 的宽与高. 可以得到不同形状的 default box, 不同长宽比的 default box 可以适应形状不同的路牌目标, 便于最终产生准确的预测框.

4) 浅层特征图与深层特征图相结合. 由于路牌数字区域在输入图像中属于小目标, 算法采用了浅层卷积特征与深层特征相融合的方式增强最终特征图中对目标区域的响应, 使得特征图中能更多的包含浅层特征中的小目标部分, 如图 4. 具体做法是将 conv5 输出的结果进行反卷积 (Deconvolution), 将特征图尺寸扩大到与 conv4 的结果一致, 将各自的 512 通道的特征图连接, 构成 1 024 维特征, 然后使用 1×1 卷积核进行卷积, 将特征图通道数减少到 512. 这样就实现了特征融合, 为后续的卷积提取 anchor box 提供了特征模板.

模型训练阶段, 通过对 VGG16 网络最后生成的融合特征模板进行多次卷积, 得到不同形状大小的 default box, 这些 default box 与真实边框 (ground truth) 匹配, 因此一个 ground truth 可能对应多个 default box. 通过对这些 default box 与真实边框的交并比 (Intersection over union, IoU) 不断优化, 得到形状与大小最接近真实值的 default box. 在算法预测阶段, 对于每个预测出的 default box, 首先根据类别置信度确定其类别与置信度值, 同时删除属于背景类别的预测框, 然后根据置信度阈值 (代码中为 0.3) 过滤掉阈值较低的预测框. 对于留下的预测框进行处理, 根据先验框得到其真实的位置参数. 处理之后, 一般需要根据置信度进行降序排列, 然后仅保留置信度最高的 k 个预测框. 最后进行非最大值抑制 (Non-maximum suppression, NMS) 算法[16], 过滤掉那些重叠度较大的预测框, 剩余的预测框就是检测结果.

在路牌检测模块训练时, 使用损失函数为

$$L(x, c, l, g) = \frac{1}{N}(\alpha L_{\text{loc}}(x, l, g) + L_{\text{conf}}(x, c)) \tag{2}$$

损失函数包括两部分, 1) 预测位置损失 $L_{\text{loc}}(x, l, g)$, α 为平衡系数; 2) 置信度损失函数 $L_{\text{conf}}(x, c)$. x 用来表示是否匹配到真实边框, g 为真实边框坐标偏移, l 与 c 分别表示预测的坐标偏移以及置信度. loss 中的 N 为与真实边框匹配的 default box 的个数, 如果没有, 则将 loss 置为 0. $L_{\text{loc}}(x, l, g)$ 是一个 $Smooth_{L1}$ loss, 即:

$$Smooth_{L1}(x) = \begin{cases} 0.5x^2, & \text{若 } |x| < 1 \\ |x| - 0.5, & \text{否则} \end{cases} \tag{3}$$

L_{loc} 的定义为

$$L_{\text{loc}}(x, l, g) = \sum_{x \in Pos}^{N} \sum_{m \in (cx, cy, w, h)} x_{ij}^k Smooth_{L1}\left(l_i^m - \hat{g}_j^m\right) \tag{4}$$

其中, $\hat{g}_j^{cx} = (g_j^{cx} - d_i^{cx})/d_i^w$, $\hat{g}_j^{cy} = (g_j^{cy} - d_i^{cy})/d_j^h$, $\hat{g}_j^w = \lg(g_j^w/d_i^w)$, $\hat{g}_i^h = \lg(g_j^h/d_j^h)$, 可以看出损失函数包含了中心坐标 cx, cy 的损失值以及边框的宽 w 和高 h 的损失量, 公式中 l 表示最终预测框的偏移, $\hat{g}^h$, $\hat{g}^w$ 表示真实边框长宽的偏移量, d^h, d^w 分别为与真实边框相匹配的 default box 的长与宽.

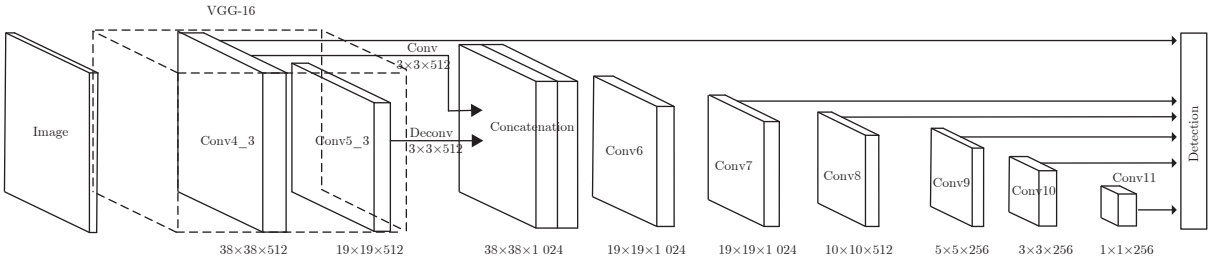


图 4　特征融合 SSD 算法结构图

Fig. 4　Feature fusion SSD algorithm structure

L_{conf} 的定义为

$$L_{\text{conf}}(x,c) = -\sum_{l\in Pos}^{N} x_{ij}^{p}\lg\hat{c}_i^{p} - \sum_{i\in Neg}\lg\left(\hat{c}_i^{0}\right) \quad (5)$$

包括正样本 (Pos) 损失与负样本 (Neg) 损失, 其中, $\hat{c}_i^p = \dfrac{\exp(c_i^p)}{\sum_p \exp(c_i^p)}$, 这是一个 softmax 函数, $\hat{c}_i^p$ 表示第 i 个 default box 是第 p 类的概率, $p=0$ 表示背景. $x_{ij}^p = 1$ 时表示第 i 个 default box 匹配到第 j 个类别为 p 的真实边框, $x_{ij}^p = 0$ 表示没有匹配到真实边框.

白天与夜间选用同样的网络结构, 但使用不同的数据集进行训练, 训练参数选择如表 1 所示.

训练模型所需要的数据全部来自视频逐帧截图, 然后再从每张图截取 512 像素 × 512 像素 大小的区域构成训练路牌检测模块的数据集, 白天数据集共包括 1 万张图片, 夜间数据集包括 4 700 张图片, 标签为 xml 格式, 如图 5 所示.

1.3 语义分割模块

语义分割模块分为白天与夜间两种情况, 其处理流程也不同.

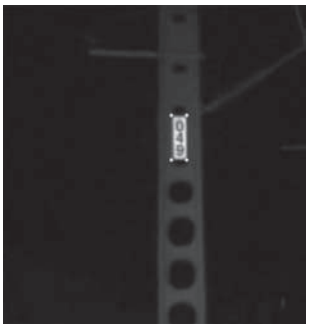
1.3.1 白天情况下的语义分割

白天情况下, 语义分割模块包括两部分, 第一部分是一个全卷积的神经网络, 此网络的作用是将得到的路牌区域去除多余的背景. 全卷积神经网络顾名思义在整个网络结构中只有卷积层和激活层, 不包含全连接层, 这样可以保持图片中物体的空间信息, 有利于实现物体的二值化. 另外将卷积层的步长设置为 1, 并且卷积时采用特征图填充的方法并且不进行池化操作, 保证图像经过每层运算后大小与原图一致. 由于输入图像的尺寸非常小, 因此选择较小的网络结构, 实验结果如表 2, 选择 6 层的网络可以很好的平衡速度与分割准确性. 网络结构图如图 6所示.

表 1 特征融合 SSD 算法训练超参数
Table 1 The hyperparameter of feature fusion SSD algorithm

超参数	取值
Iteration	5 000
Learn rate	0.00099
Learn rate decay factor	0.98
Weight decay	0.0005
Batch size	16
Optimizer	SGD

(a) 白天条件下的示例图片
(a) Example image from day



(b) 夜间条件下的示例图片
(b) Example image from night

图 5 路牌检测模块的训练数据集示例图
Fig. 5 Example image of the training data set of the road sign detection module

表 2 不同结构下分割准确率及速度
Table 2 Segmentation accuracy and speed in different structures

网络层数	重叠率 (%)	速度 (s/张)
5	81.15	0.006
6	82.19	0.010
7	81.38	0.017
8	81.8	0.019
9	82.02	0.020

在 6 个卷积层中, 卷积核尺寸选择如表 3 所示.

由于前一模块输出的区域图片大小不一, 因此在送入神经网络之前需要对输入图像四周进行填充, 填充像素值为 255, 填充后图像分辨率为 72 像素 × 72 像素. 模块第二部分的功能是单个数字分割, 目标图像经过语义分割网络处理后, 图像中像素值为 0 或 255, 如图 7(a); 通过对图像进行形态学操作 (膨胀与腐蚀), 消除噪声区域, 如图 7(b) 和 7(c); 之后通过轮廓搜索找到数字所在的矩形框如图 7(d); 将矩形框进行三等分操作, 即可得出分离的数字, 如图 7(e).

语义分割网络训练时损失函数采用交叉熵损失

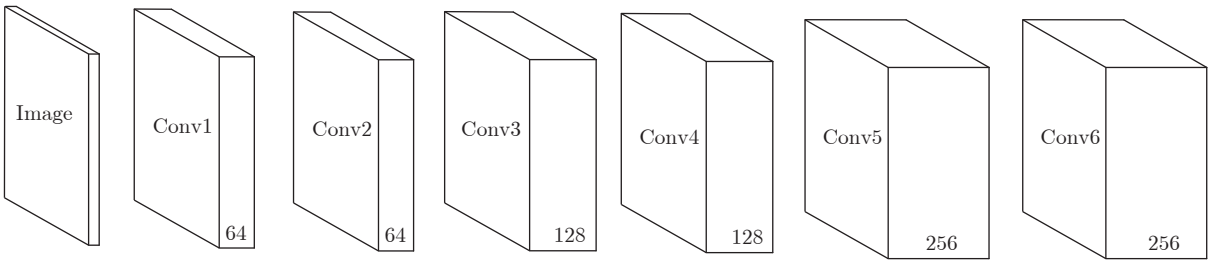


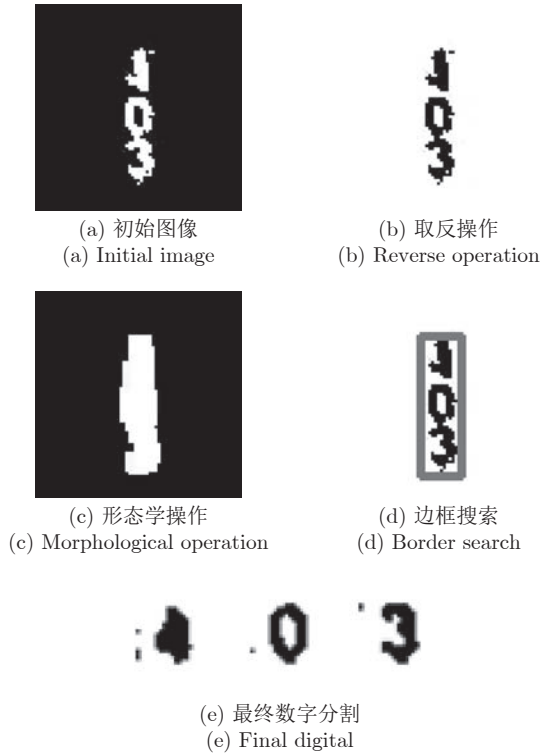
图 6　语义分割网络结构图

Fig. 6　Semantic segmentation network structure

表 3　语义分割网络卷积核大小

Table 3　The kernel size of semantic segmentation network convolution

网络层	卷积核大小	输出大小
Input	–	72 × 72
Conv1	3×3	72 × 72 × 64
Conv2	3×3	72 × 72 × 64
Conv3	3×3	72 × 72 × 128
Conv4	3×3	72 × 72 × 128
Conv5	3×3	72 × 72 × 256
Conv6	3×3	72 × 72 × 256



(a) 初始图像　(b) 取反操作

(a) Initial image　(b) Reverse operation

(c) 形态学操作　(d) 边框搜索

(c) Morphological operation　(d) Border search

(e) 最终数字分割

(e) Final digital

图 7　白天条件下语义分割模块单独数字分割过程

Fig. 7　Semantic segmentation module separate digital segmentation process under daytime conditions

函数. 交叉熵刻画的是实际输出 (概率) 与期望输出 (概率) 的距离, 也就是交叉熵的值越小, 两个概率分布就越接近. 交叉熵公式为

$$H(P,Q) = -\sum_{x}[(P(x)\lg Q(x) + (1-P(x))(\lg(1-Q(x)))] \tag{6}$$

其中, P 为期望输出, Q 为实际输出, 交叉熵值越小, 两个概率分布越接近. 在算法实际应用中, 输出的分割图像经过 softmax 函数将每个点像素值转换为 0 ∼ 1 之间的概率, 式中的 Q 对应的标注图像中的数字部分为 1, 背景部分为 0, 对应 P. 训练超参数选择如表 4 所示.

表 4　语义分割算法训练超参数

Table 4　The hyperparameter semantic segmentation algorithm

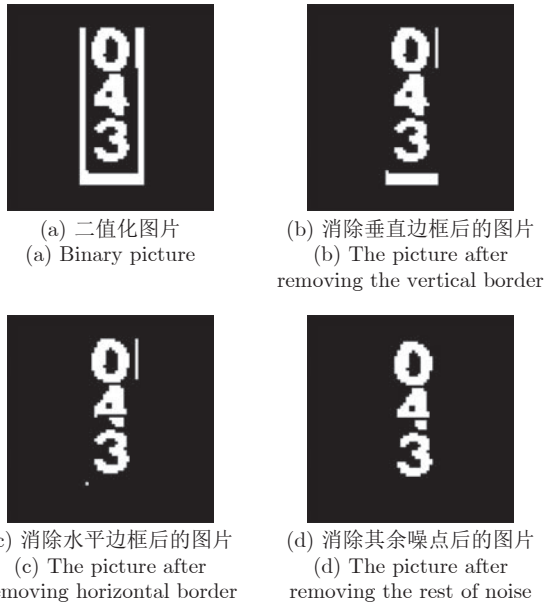
超参数	取值
Iteration	100 000
Learn rate	0.0001
Learn rate decay factor	0.98
Batch size	2
Optimizer	SGD

训练分割模块的数据集来目标检测网络输出的检测结果共包括 9 856 张图片, 其中 9 280 张用于训练 (8 628 张白天图像, 652 张夜间图像), 576 张用于测试 (483 张白天图像, 93 张夜间图像), 真实值由以上图片二值化后再精细处理, 得到没有噪点的数字区域.

1.3.2　夜间情况下的语义分割

在夜间情况下, 与白天的场景差异比较大, 通过对夜间视频分析可知, 视频中高亮度区域为列车车灯照射的白色路牌以及其中的数字, 远处背景则为黑色, 目标区域与背景有较大区别, 如图 5(b) 所示. 因此可以在目标检测算法后直接进行形态学处理, 得到数字区域的二值图像, 具体操作如下: 1) 对

路牌检测模块所得到的路牌区域进行预处理, 得到含有部分路牌边框以及噪点的二值图像, 将其记为 A, 如图 8(a). 2) 通过形状为矩形, 尺寸为 1×25 的结构元素, 对 A 进行形态学开运算 (先腐蚀后膨胀), 将得到的结果图像记为 B, 然后通过 $A-B$ 得到消除垂直状背景的图像, 将其记为 C, 如图 8(b). 3) 通过形状为矩形, 尺寸为 12×1 的结构元素, 同样对 A 进行开运算得到图像, 将其记为 D, 通过 $C-D$ 得到消除了边框区域的图像 E, 如图 8(c). 4) 通过尺寸为 2×2 的结构元素, 对图像 E 进行开运算后, 消除其他噪点, 得到最终的二值图像, 如图 8(d).



(a) 二值化图片
(a) Binary picture

(b) 消除垂直边框后的图片
(b) The picture after removing the vertical border

(c) 消除水平边框后的图片
(c) The picture after removing horizontal border

(d) 消除其余噪点后的图片
(d) The picture after removing the rest of noise

图 8 夜间条件下路牌中数字区域的二值化过程

Fig. 8 Binary process of digital regions in rail signs under night conditions

1.4 数字识别模块

数字识别模块由一个小型的卷积神经网络构成, 首先将接收到的分割好的单独数字图片进行预处理, 具体操作是将图片进行通道转换, 将三通道 RGB 图像转换为灰度图像, 然后进行直方图均衡, 消除一部分光照不均的产生的影响, 然后将灰度图像转换为二值图像, 并在周围填充白色区域, 统一图像大小. 最后将图像像素值归一化至 [0,1] 区间, 通过卷积层提取特征后送入全连接层, 最后输出识别结果. 由于路牌数字都为印刷体数字, 因此白天与黑天条件可以使用同一网络进行检测. 模块网络结构如图 9 所示.

考虑到输入待检测图片尺寸比较小, 因此选择构建了一个小型的卷积神经网络, 数字识别网络由三个卷积层, 两个池化层以及三个全连接层组成. 由于网络中包含两个池化层, 导致输出特征图大小会减少到原图的 1/4, 因此将输入图像分辨率定为为 28×28, 防止由于池化导致后续的特征图分辨率过低, 从而降低模型的识别效果. 算法训练超参数选择如表 5 所示.

训练数据集共包括 5 703 张 0～9 的数字二值化图片, 全部为印刷体.

1.5 后处理

通过以上三个模型, 可以将每一帧的图片中路牌的数字识别出来, 但是由于实际输入是视频, 因此同一个路牌会出现在视频的多个帧中, 所以可以通过对这多个帧进行检测, 输出这些帧中出现次数最多的检测结果, 此外, 由于路牌数字是前后具有关联性, 比如后一个路牌都比前一个数字大 2, 所以可以通过前后路牌的数值关系进行结果修正. 具体操作如下:

1) 利用同一路牌不同帧之间的位置关系. 当前帧中的路牌为 x_i, 记录路牌的左上角坐标为 (a_i, b_i), 由于列车向前行驶, 视频中的同一个路牌由远及近路牌位置不断向当前帧的左上方移动, 如果 $a_{i+1} < a_i$ 且 $b_{i+1} < b_i$, 那么可以确定下一帧中的路牌 x_{i+1} 与

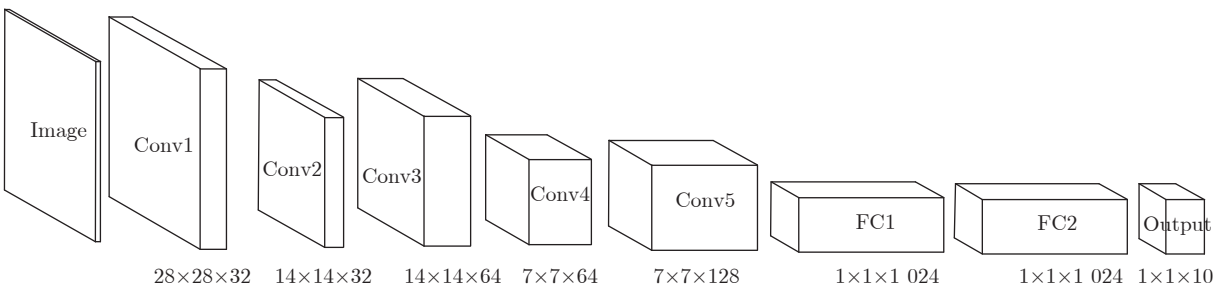


图 9 数字识别网络结构图

Fig. 9 Digital identification network structure

表 5　数字识别算法训练超参数
Table 5　The hyperparameter of digital recognition algorithm

超参数	取值
Iteration	80 000
Learn rate	0.001
Learn rate decay factor	0.99
Batch size	8
Optimizer	SGD

x_i 中的路牌为同一个路牌, 将x_{i+1} 计入集合 $X=\{x_i, x_{i+1}\}$, 继续执行直到条件不满足, 得到集合 $X=\{x_i, x_{i+1}, \cdots, x_{i+n}\}$, 计算集合中出现最多的数字即为这 n 帧中路牌的数字 o_x, 同时将数字计入集合 O.

2) 利用不同路牌数字大小之间的逻辑关系. 将不满足步骤 1) 中条件的帧记为 y_i, 按步骤 1) 中的规则得到 $Y=\{y_i, y_{i+1}, \cdots, y_{i+n}\}$, Y 中出现次数最多的数字记为 o_y, 判断 o_y 与 o_x 之间的差值是否为 2, 若为 2, 则这 n 帧中路牌数字为 o_y, 否则判断 $o_x \pm 2$ 是否在集合 Y 中出现过, 如果出现过, 则 $o_y = o_x \pm 2$.

通过前后路牌的逻辑关系修正最终输出结果, 对最终系统的准确率有很可观的提升.

2　实验

算法训练在服务器中进行, 使用语言为 python2.7, 使用深度学习框架为 tensorflow-GPU (1.12.0), CUDA (9.0.176), CUDNN (7.1.4), 服务器配置如表 6 所示.

2.1　路牌检测实验

首先验证本文提出的特征融合 SSD 与直接使用浅层特征 SSD, 哪个检测小目标 (路牌) 更优. 对比效果如表 7 所示, 可以看出直接使用浅层特征进行目标检测, 其 mAP 不如特征融合, 同时也验证了第 1.2 节中的理论分析, 即:浅层特征感受野较小, 容易包含小目标特征, 但是不足以描述目标类别, 而深层特征虽然可以区分目标类别, 但是由于高度语义化, 损失了位置信息, 对小目标的位置描述不足, 因此需要将浅层特征与深层特征融合.

表 7　不同网络结构下的测试结果
Table 7　Test results in different training datasets

网络结构	测试集	mAP (%)
特征融合 SSD	白天 + 黑天	80.10
浅层特征 SSD (VGG-2)	白天 + 黑天	76.23
浅层特征 SSD (VGG-3)	白天 + 黑天	79.01
浅层特征 SSD (VGG-4)	白天 + 黑天	79.89

视频分为白天与黑天两种情况, 因此分别在白天数据集、黑天数据集以及两者合并的数据集分别进行网络训练, 得到表 8 中结果. 由表 8 可知在训练路牌检测模型的时候, 将训练集分为白天、黑天两类进行训练效果更好, 从而分别得到适用于白天情况的路牌检测模型以及适用于黑天情况的路牌检测模型.

路牌检测算法损失函数包括位置损失函数 L_{loc} 与置信度损失函数 $L_{\text{conf\_Neg}} + L_{\text{conf\_Pos}}$, 经过 8 万轮训练, 在白天条件下, 路牌检测模块训练集损失函数变化如图 10 所示. 由图 10 可知, 算法在训练过程中对于真实目标始终有预测框与之匹配, 因此正样本 (Pos) 的损失值 (图 10(a)) 保持在 2 左右, 而背景即负样本 (Neg) 的损失值 (图 10(b)) 不断下降, 说明算法在不断地学习中逐渐排除了背景的干扰, 这样使得算法可以更准确地识别路牌, 并且随着位置损失值 (Localization) 不断下降 (图 10(c)), 定位也越来越准确, 算法的总损失函数值 (图 10(d)) 也不断减小并趋于稳定. 经过测试, 平均精度为 81.29 %.在黑天条件下, 采用的特征融合的 SSD 算法训练 6 万次, 路牌检测模块损失函数变化如图 11 所示. 由于路牌检测模块在白天和黑天两种条件下, 使用的深度学习模型框架一样, 只是训练集有所不同, 因此损失函数变化大致相同. 经过测试集测试平均精度为 83.57 %. 白天条件下, 路牌检测模块测试效果图如图 12所示; 黑天条件下, 路牌检测效果如图 13 所示.

表 6　服务器配置
Table 6　Server configuration

设备	配置
系统	Ubuntu16.04
内存	64 GB
GPU	Tesla K40
显存	12 GB

表 8　不同训练集下的测试结果
Table 8　Test results in different training datasets

训练集	测试集	算法	mAP (%)
白天	白天	特征融合 SSD	81.29
黑天	黑天	特征融合 SSD	83.57
白天 + 黑天	白天	特征融合 SSD	80.12
白天 + 黑天	黑天	特征融合 SSD	79.30

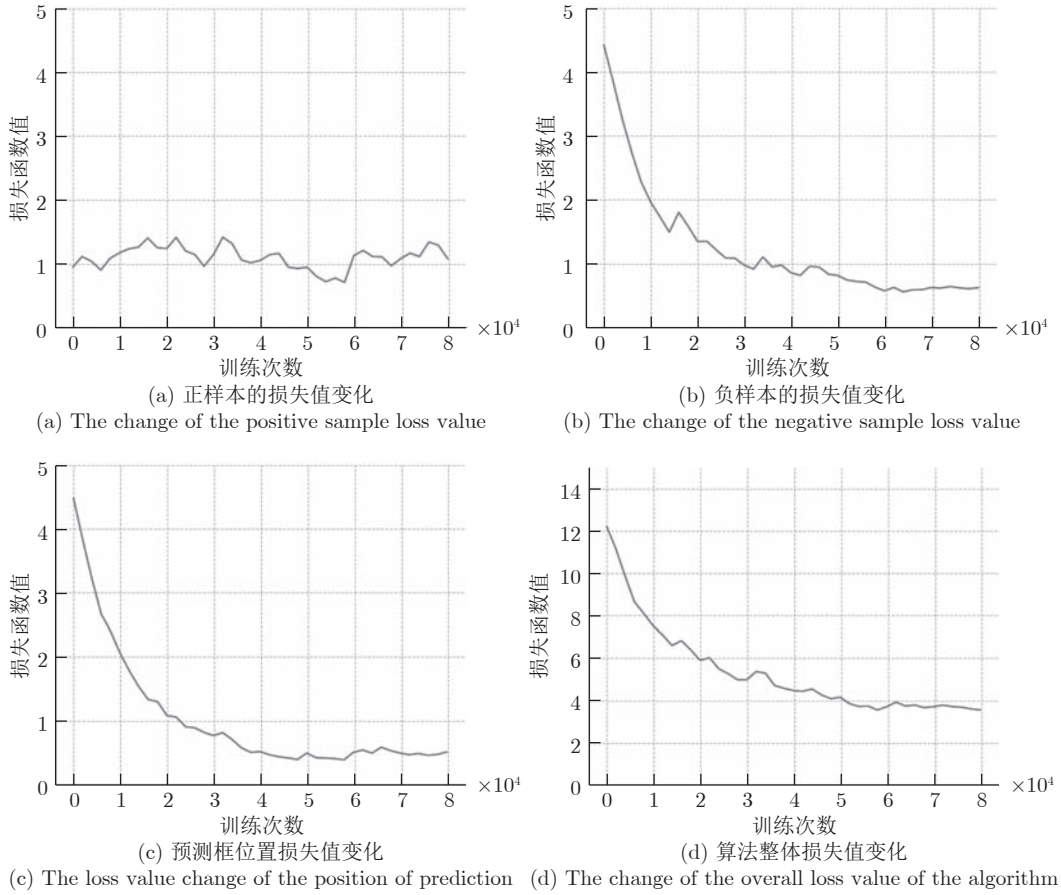


图 10 白天条件下路牌检测模块训练损失值变化图

Fig. 10 The rail sign detection module training loss value change graph under the daytime conditions

2.2 语义分割实验

语义分割网络的训练次数为 15 万次, 目的是将四周填充过白色区域的路牌区域图片进行分割, 得到只含有数字区域的二值化图片, 消除背景与噪声的干扰. 之后通过轮廓搜索, 得到精确的数字区域位置并分割出单独数字.

为了验证训练语义分割模型时, 使用哪种数据集更好, 本实验对白天、黑天两种情况分别进行测试, 测试结果如表 9 所示. 通过对比可知, 处理白天数据时, 使用深度学习的方法在 “白天+黑天” 的训练集来训练语义分割模型. 处理黑天数据时, 使用 “阈值分割 + 形态学” 在准确率和速度上两个方面, 均优于使用深度学习的结果, 因此在黑天情况下, 本文使用 “阈值分割 + 形态学” 的方法来处理.

训练过程中, 语义分割模块训练集损失函数的变化如图 14 所示, 损失函数随着迭代次数的增加而减小, 在 15 万次迭代后, 达到稳定值 0.014 左右. 经过测试集测试, 分割准确率为 82.19 %.

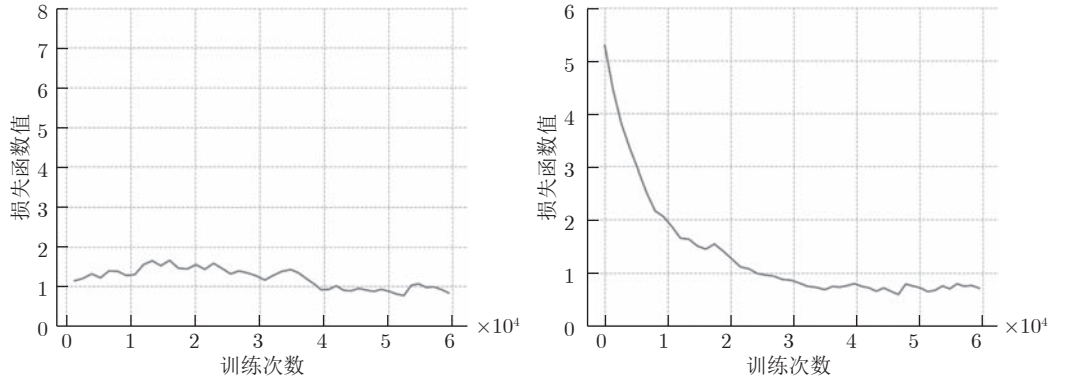
语义分割的结果如图 15 所示, 第 1 列为原图, 第 2 列为真实值, 第 3 列为语义分割模型的预测值. 通过对比真实值与预测值, 可知本文语义分割模型可以较好地在路牌区域内进一步得到数字区域. 该数字区域中, 共包含三个数字, 将其分别拆分成单个数字区域, 其结果如图 16 所示.

2.3 数字识别实验

数字识别模块对单个数字区域进行分类, 该模块训练集损失函数变化值如图 17 所示.

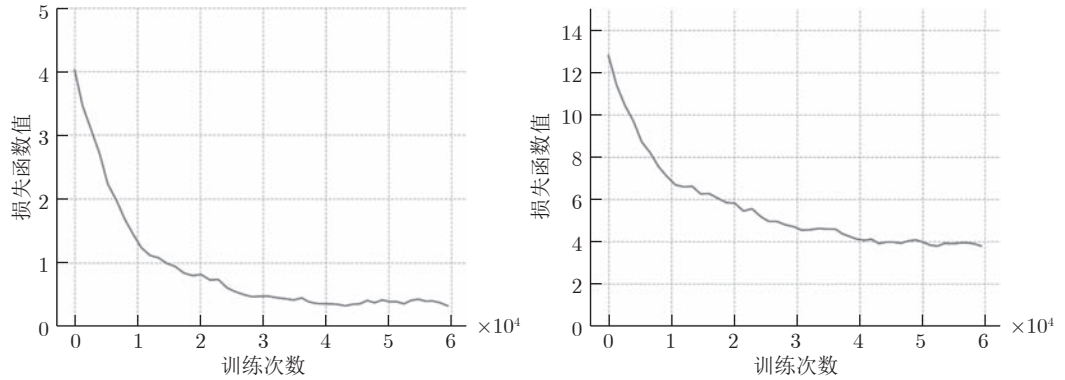
由于输入图像为二值图像, 因此可以将白天与黑天下产生的二值图像放在一个训练集中进行训练, 这样得到的模型可以适用白天与黑天两种条件训练时加入 dropout 层, 系数为 0.8, 防止网络过拟合, 训练完成后在测试集 305 张图片上进行测试, 准确率为 93.44 %, 符合数字识别所需要的准确率.

在所有模块训练完成后, 将三个模块整合在一起, 得到完整的系统, 并且加入前后路牌间位置和数字的逻辑关系这个两个先验知识, 从而消除检测错误的区域, 保证每张图中最多只有一个路牌被检测到, 经过测试, 系统的最终准确率如表 8 所示, 系



(a) 正样本的损失值变化
(a) The change of the positive sample loss value

(b) 负样本的损失值变化
(b) The change of the negative sample loss value



(c) 预测框位置损失值变化
(c) The loss value change of the position of prediction

(d) 算法整体损失值变化
(d) The change of the overall loss value of the algorithm

图 11　黑天条件下路牌检测模块训练损失值变化图

Fig. 11　The rail sign detection module training loss value change graph under the dark conditions

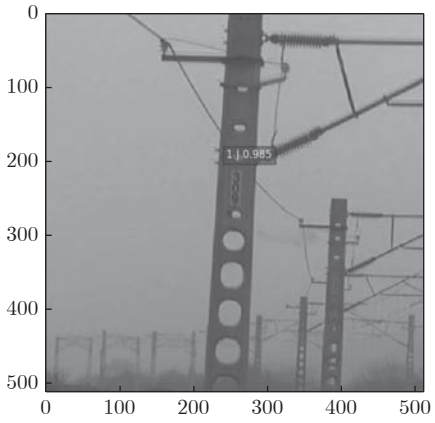


图 12　白天路牌检测模块效果图

Fig. 12　The result of the daytime rail sign detection module

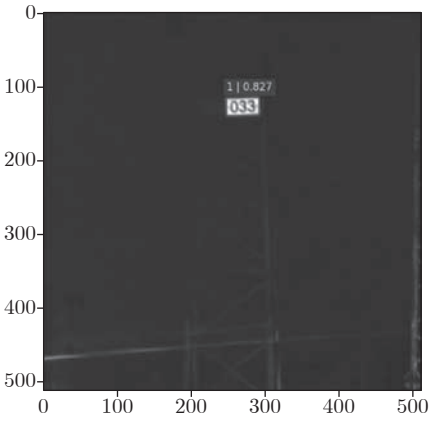


图 13　黑天路牌检测模块效果图

Fig. 13　The result of the night rail sign detection module

统的整体效果运行如图 18 所示. 黑天情况下虽然对比度更加明显, 但是实际视频中有些路牌由于列车灯光照射而过度曝光, 导致路牌中数字信息丢失, 导致数字识别模块中无法对路牌中的数字进行识别, 所以黑天情况下识别准确率低于白天.

由表 10 可知, 路牌检测平均每帧用时 0.07 s,

表 9 不同训练集与分割方法下的分割效果
Table 9 Segmentation effects under different training sets and segmentation methods

使用方法	训练集	测试集	mAP (%)	速度 (s/张)
深度学习	白天+黑天	白天	83.67	0.001
深度学习	白天	白天	82.19	0.001
深度学习	白天+黑天	黑天	83.07	0.001
阈值分割+形态学	黑天	黑天	85.39	0.0001

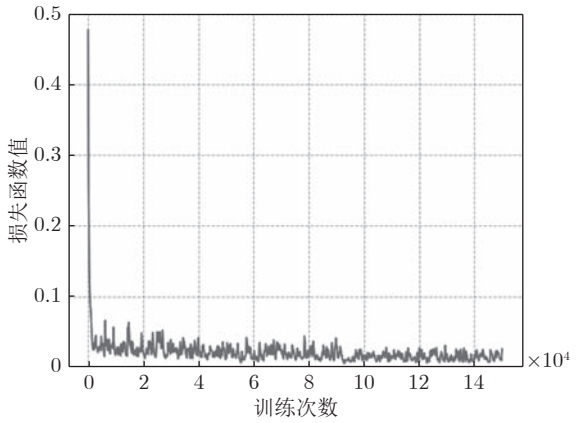
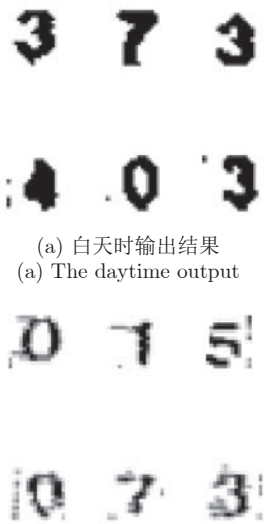


图 14 语义分割网络损失函数变化图
Fig. 14 Semantic segmentation network loss function curve in the training

图 15 语义分割网络实验效果图
Fig. 15 Semantic segmentation network experiment results

语义分割平均每帧用时 0.01 s, 数字识别平均每帧用时 0.005 s, 黑天时数字分割平均每帧用时 0.001 s. 因此总体来说, 白天条件下, 处理每帧图像平均用时 0.085 s, 相当于 11.8 帧/s; 黑天条件下, 处理每帧图像平均用时 0.068 s, 相当于 14.7 帧/s.



(a) 白天时输出结果
(a) The daytime output

(b) 夜间时输出结果
(b) Output during night

图 16 单个数字区域的结果图
Fig. 16 Regions of each digital number obtained by semantic segmentation algorithm

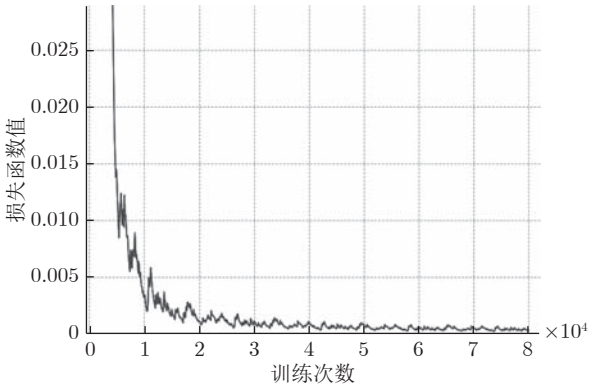
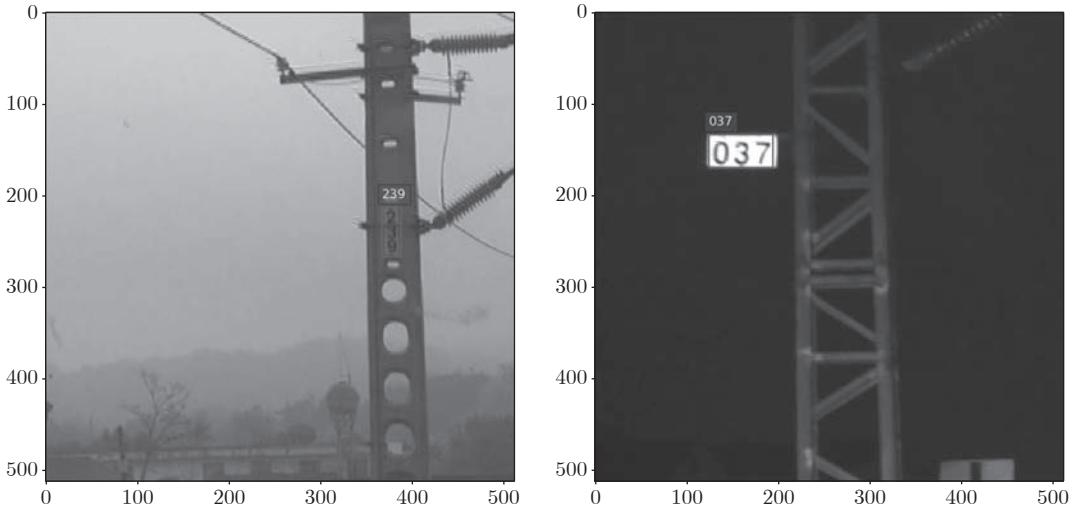


图 17 数字识别网络损失函数变化图
Fig. 17 Digital recognition network loss function curve in the training

3 结论

本文基于卷积神经网络, 面向智能轨道交通, 提出一个针对高铁运行过程中铁路两侧的路牌识别模型. 该模型由路牌区域检测模块、语义分割模块、数字识别模块组成, 从而实现对视频中每一帧图片进行检测、分割、分类、识别, 并对传统 SSD 模型进行改进, 实现小目标的识别, 对传统语义分割模型进行改进, 实现对小尺寸图像的精确分割. 实验表明, 本文方法能够以白天 11.8 帧/s、黑天 14.7 帧/s 的速度对铁路两侧的路牌数字进行识别, 白天条件下的平均识别准确率为 87.98 %, 黑天条件下的平



(a) 白天条件下的系统整体运行效果
(a) The daytime conditions

(b) 黑天条件下的系统整体运行
(b) The night conditions

图 18 系统整体运行效果图

Fig. 18 Total operation of the system

表 10 系统准确率统计 (%)

Table 10 Statistics of each system accuracy rate (%)

使用方法	准确率
目标检测 + 语义分割 + 数字识别 (白天)	73
目标检测 + 语义分割 + 数字识别 + 后处理 (位置关系) (白天)	78
目标检测 + 语义分割 + 数字识别 + 后处理 (位置关系 + 逻辑关系) (白天)	87.98
目标检测 + 语义分割 + 数字识别 (黑天)	58.33
目标检测 + 语义分割 + 数字识别 + 后处理 (位置关系 + 逻辑关系) (黑天)	72.92

均识别准确率为 72.92 %.

References

1 Li Ze-Xin. Application of artificial intelligence in intelligent transportation. *Public Communication of Science and Technology*, 2018, **10**(19): 104−105
(李泽新. 人工智能在智能交通中的应用. 科技传播, 2018, **10**(19): 104−105)

2 Feng Jiang-Hua. Technical evolution and intelligent development of rail transit equipments. *Control and Information Technology*, 2019, **1**(1): 1−6
(冯江华. 轨道交通装备技术演进与智能化发展. 控制与信息技术, 2019, **1**(1): 1−6)

3 Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA: IEEE, 2014. 580−587

4 Uijlings J, van de Sande K, Gevers T, Smeulders A. Selective search for object recognition. *International Journal of Computer Vision*, 2013, **104**(2): 154−171

5 He K M, Zhang X Y, Ren S Q, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, **37**(9): 1904−1916

6 Girshick R. Fast R-CNN. In: Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Chile: IEEE, 2015. 1440−1448

7 Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, Nevada, USA: IEEE, 2016. 779−788

8 Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE, 2017. 6517−6525

9 Viola P, Jones M. Object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, USA: IEEE, 2001. 1−1

10 Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. San-Diego, USA: IEEE, 2005. 886−893

11 Girshick R, McAllester D, Ramanan D. Object detection with discriminatively trained part —— based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627−1645

12 Shelhamer E, Long J, Darrell T. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, **39**(4): 640−651

13 Liu W, Dragomir A, Dumitru E, Christian S, Scott R, Fu C Y, et al. SSD: single shot multiBox detector. In: Proceedings of the 2016 European Conference on Computer Vision. Amsterdam, The Netherlands: ECCV, 2016. 21−37

14 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: Proceedings of the 2015 International Conference on Learning Representations. San Diego, USA: ICLR, 2015.

15 Ren S Q, He K M, Girshick R, Sun J. Faster R-CNN: towards

real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, **39**(6): 1137−1149

16 Neubeck A, Van G. Efficient non maximum suppression. In: Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China: IEEE, 2006. 850−855

孟 琭 东北大学信息科学与工程学院副教授. 2010 年获东北大学博士学位. 主要研究方向为人工智能, 图像处理.
E-mail: menglu@ise.neu.edu.cn
(**MENG Lu** Associate professor at the College of Information Science and Engineering, Northeastern University. He received his Ph.D. degree from Northeastern University in 2010. His research interest covers artificial intelligence and image processing.)

孙霄宇 东北大学信息科学与工程学院硕士研究生. 2017 年获武汉科技大学学士学位, 2020 年获东北大学硕士学位. 主要研究方向为图像处理, 目标检测.
E-mail: ovxex081@163.com
(**SUN Xiao-Yu** Master student at the College of Information Science and Engineering, Northeastern University. He received his bachelor degree from Wuhan University of Science and Technology in 2017 and the master degree from Northeastern University in 2020, respectively. His research interest covers image processing and object detection.)

赵 滨 友和利德科技有限公司创始人兼 CEO. 2014 年毕业于美国德克萨斯大学阿灵顿分校工商管理系.主要研究方向为数据分析.
E-mail: zhaob@uuvalue.com
(**ZHAO Bin** Founder and CEO of UUValue Technology Co., Ltd. He graduated from the University of Texas at Arlington, USA, in 2014. His main research interest is data analysis.)

李 楠 沈阳产品质量监督检验院高级工程师. 2015 年获得东北大学检测技术与自动化装置博士学位. 主要研究方向为检测技术.
E-mail: 875875@163.com
(**LI Nan** Senior engineer of Shenyang Product Quality Supervision and Inspection Institute. He received his Ph.D. degree in testing technology and automation equipment from Northeastern University in 2015. His main research interest is detection technology.)