

基于贝叶斯 CNN 和注意力网络的钢轨表面缺陷检测系统

金侠挺^{1,2} 王耀南^{1,2} 张辉^{2,3} 刘理^{1,2} 钟杭^{1,2} 贺振东⁴

摘要 面向复杂多样的钢轨场景, 本文扩展了最先进的深度学习语义分割框架 DeepLab v3+ 到一个新的轻量级、可伸缩性的贝叶斯版本 DeeperLab, 实现表面缺陷的概率分割. 具体地, Dropout 被融入改进的 Xception 网络, 使得从后验分布中生成蒙特卡罗样本; 其次, 提出多尺度多速率的空洞空间金字塔池化 (Atrous spatial pyramid pooling, ASPP) 模块, 提取任意分辨率下的密集特征图谱; 更简单有效的解码器细化目标的边界, 计算 Softmax 概率的均值和方差作为分割预测和不确定性. 为解决类别不平衡问题, 基于在线前景 - 背景挖掘思想, 提出损失注意力网络 (Loss attention network, LAN) 定位缺陷以计算惩罚系数, 从而补偿和抑制 DeeperLab 的前景与背景损失, 实现辅助监督训练. 实验结果表明本文算法具有 91.46% 分割精度和 0.18 s/帧的运行速度, 相比其他方法更加快速鲁棒.

关键词 钢轨表面缺陷, 视觉检测, 贝叶斯卷积神经网络, 注意力机制, 类别不平衡

引用格式 金侠挺, 王耀南, 张辉, 刘理, 钟杭, 贺振东. 基于贝叶斯 CNN 和注意力网络的钢轨表面缺陷检测系统. 自动化学报, 2019, 45(12): 2312-2327

DOI 10.16383/j.aas.c190143

DeepRail: Automatic Visual Detection System for Railway Surface Defect Using Bayesian CNN and Attention Network

JIN Xia-Ting^{1,2} WANG Yao-Nan^{1,2} ZHANG Hui^{2,3} LIU Li^{1,2} ZHONG Hang^{1,2} HE Zhen-Dong⁴

Abstract This paper extends the state-of-the-art deep learning framework DeepLab v3+ to a light-weighted and scalable Bayesian version DeeperLab for the defect detection on complex and diverse rail surface. Specifically, Dropout is incorporated into the improved Xception network for Monte Carlo sampling from posterior distribution. Atrous spatial pyramid pooling (ASPP) module is utilized to extract the dense features at multiple scales and rates. Furthermore, a simpler and efficient decoder is proposed to improve the defect edges, and outputs the mean and variance of Softmax probability as segmentation and uncertainty. To solve class imbalance problem, we present the loss attention network (LAN) to perform auxiliary supervision for DeeperLab training. Experimental results show that the proposed algorithm is more accurate and robust than other methods with 91.46% precision and 0.18 s/frame speed.

Key words Rail surface defects, visual detection, Bayesian convolutional neural network (CNN), attention mechanism, class imbalance

Citation Jin Xia-Ting, Wang Yao-Nan, Zhang Hui, Liu Li, Zhong Hang, He Zhen-Dong. DeepRail: automatic visual detection system for railway surface defect using Bayesian CNN and attention network. *Acta Automatica Sinica*, 2019, 45(12): 2312-2327

收稿日期 2019-03-07 录用日期 2019-08-08

Manuscript received March 7, 2019; accepted August 8, 2019

国家自然科学基金 (61573134, 61733004), 湖南省科技计划项目 (2017XK2102, 2018GK2022, 2018JJ3079) 资助

Supported by National Natural Science Foundation of China (61573134, 61733004) and Hunan Key Project of Research and Development Plan (2017XK2102, 2018GK2022, 2018JJ3079)

本文责任编辑 阳春华

Recommended by Associate Editor YANG Chun-Hua

1. 湖南大学电气与信息工程学院 长沙 410082 2. 湖南大学机器人视觉感知与控制技术国家工程实验室 长沙 410082 3. 长沙理工大学电气与信息工程学院 长沙 410114 4. 郑州轻工业大学电气与信息工程学院 郑州 450000

1. College of Electrical and Information Engineering, Hunan University, Changsha 410082 2. National Engineering Laboratory of Robot Vision Perception and Control Technology, Hunan University, Changsha 410082 3. College of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114 4. College of Electrical and Information

铁路元件故障诊断是维系列车安全运行的重要保障, 轨道表面语义分割要求了像素级别的场景解析. 处于服役状态的钢轨受到轮轨的挤压、冲击、磨损等接触应力以及材料微观组织退化等影响, 其健康状况和性能质量不断恶化, 从而形成塌陷、剥离、断轨、滚动接触疲劳裂纹、内核损伤、压溃、侧磨、腐蚀等表面缺陷. 例如, 塌陷是由于车轮撞击轨道而导致“擦伤”的金属疲劳, 并随着时间推移使得钢轨退化. 母材从轨头脱落的现象称为剥离, 如果不加以控制, 这些潜在损伤将进一步恶化金属而造成断轨. 可见, 伤损模式与缺陷类型的多样性和动态

Engineering, Zhengzhou University of Light Industry, Zhengzhou 450000

性给已有的轨道检查技术带来了巨大挑战。

传统的视觉方法^[1-5]利用低级别的灰度特征, 应用于无标记的钢轨样本. 贺振东等^[1]提出反向 P-M 扩散模型进行图像增强和提取缺陷边缘. Re-sendiz 等^[3]结合 Gabor 滤波器与光谱估计, 提出多信号分类 (Multiple signal classification, MUSIC) 算法以识别轨道元件、道岔和结缘节. Liang 等^[5]提出基于 ANC 和时频特性的方法, 识别车轮扁疤和钢轨表面缺陷. 但是, 这些方法都不适用于基于条件的复杂钢轨场景, 例如低信噪比、反射不均、铁锈油污、光照不足和过度曝光.

1) 基于深度卷积神经网络的缺陷检测

近年来, 很多成功的深度学习框架相继提出用于各类铁路元件检测, 包括紧固件^[6-7]、轨面塌陷及裂缝^[8-9]、松散股^[10-11]和开口销^[12]. Gibert 等^[6-7]提出多任务深度架构来检测铁路轨道的拉杆和紧固件. Faghih-Roohi 等^[8]提出了具有多种结构和激活性能的深度卷积神经网络 (Deep convolutional neural networks, DCNN), 用于钢轨缺陷检测. Masci 等^[9]提出了一种最大池化 CNN 对钢材缺陷分类, 避免了耗时的预处理. Chen 等^[10]级联了三个基于 DCNN 的检测阶段, 从粗到细地定位和识别悬臂节点及其紧固件缺陷. Liu 等^[11]在 Faster R-CNN 中融入特征提取模块 ILNET, 利用基于马尔科夫随机场的分割方法 ICM (Intersecting cortical model)/MPM (Maximization of the posterior marginal), 定位和分割等电线路的松散股. Zhong 等^[12]研究一种改进的三阶段缺陷检测网络 PVANET++, 用于开口销故障诊断.

然而, 大多数深度学习方法产生点估计作为输出, 因此无法获得模型对其预测的置信度. 随着 AI 系统在真实场景的应用, 比如自动驾驶、医疗诊断和工业检测, 模型置信度的额外知识对于决策是非常重要的, 特别在主动学习^[13]、半监督学习^[14]和标签传播^[15].

2) 贝叶斯深度学习

建立模型不确定性的神经网络为贝叶斯神经网络^[16-17]. 通过推断网络权重的分布, 提供了深度学习模型的概率解释. 然而, 推理大型网络的后验分布是非常困难的, 通常在不显著提高模型性能的情况下增加了参数量, 导致昂贵的计算成本.

另一方面, 训练数据的严重缺乏极有可能导致参数化卷积网络的过度拟合. 因此, 正则化方法 Dropout^[18]广泛用于防止特征过拟合和自适应. 在随机梯度下降训练中, Dropout 随机移除网络单元来从更多宽度缩减的稀疏网络中采样. 在测试时, 标准 Dropout 则近似于加权平均的效果, 即使用非

稀疏网络的权重对所有稀疏网络的预测进行均衡化.

文献 [19] 中, Dropout 作为网络权值的近似贝叶斯的变分推理法^[20]. 具体地, 在测试时使用 Dropout 随机采样网络单元, 等效于对卷积滤波器的权值施加一个伯努利分布, 而不需要任何额外的模型参数. 因此, 作者认为这是从模型的后验分布中获得的蒙特卡罗样本, 从而样本的均值可解释为预测输出, 而方差可解释为模型不确定性. 本文正是利用该技术搭建了像素级别的概率语义分割.

3) 注意力机制

为了强化模型性能, 将最新的注意力机制引入到深度学习以构成 “CNN + Attention” 体系. 一方面, 它们用于学习权重分布, 即对输入数据或特征图的不同部位的专注度不同. Fu 等^[21]训练一个注意力提议网络 APN, 将精细尺度只聚焦在原图的显著区域. Wang 等^[22]提出增强特征图的感受野, 突出有利于分类的特征, 相当于增大了网络深度.

另一方面, 注意力机制用于任务解耦. He 等^[23]提出 Mask RCNN 实例分割, 重新分配了分类和分割网络的学习能力, 极大降低了原始任务的难度. Lin 等^[24]对浅层网络的 Loss 进行引导, 只反传粗略目标位置上的 Loss 而忽略其他对象, 从而加快网络训练和降低学习难度. 本文中, 提出了一个注意力网络来补偿和抑制另一网络中不同类别的 Loss 分量, 实现辅助训练监督. 综上所述, 本文提出一种基于多模型的钢轨表面缺陷检测系统, 称作 DeepRail. 主要贡献如下:

- 1) 提出一种基于多模型的钢轨表面缺陷检测系统 DeepRail, 用于铁路轨道的精确鲁棒检测.
- 2) 一个新的轻量级、可伸缩性的贝叶斯卷积神经网络 DeeperLab 获取概率分割和模型不确定性.
- 3) 改进 Xception 网络融入 Dropout 来近似后验分布, 通过从网络权值的伯努利分布中采样.
- 4) 提出注意力网络 (Loss attention network, LAN) 监督训练 DeeperLab, 降低原始任务的难度, 达到最优权重.

1 系统概述

1.1 钢轨表面成像系统

真实工况中, 钢轨场景的成像质量受到诸多因素影响, 例如光强、曝光时间、入射角、相机运动速度、线速和振动等. 因此, 特别设计了一种成像机制将用于图像信息的获取. 该成像系统安装于一辆特别设计的小车上, 其三维示意图如图 1 所示. 核心组件是 DalsaSpyder3 行扫描相机, 分辨率为

1 024 像素, 最大行速率为 68 000 行/s, 可以获得一系列连续的视频帧. 相机速度与线率的不匹配将导致图像失真. 为了同步采集图像, 线扫描相机由轮式编码器触发, 弧槽调整入射角, 光控制器调节光强和曝光时间, 全景图是基于组帧生成的, 振动槽通过阻尼器减少振动. 此外, 自然光与污迹、铁锈等因素导致了复杂钢轨场景, 给表面缺陷检测带来挑战性.

1.2 钢轨表面检测算法

如图 2 所示, 面向复杂多样的钢轨表面场景, 本文扩展了最先进的深度学习语义分割框架 Deeplab v3+ 到一个新的轻量级、可伸缩性的贝叶斯版本 DeeperLab, 实现表面缺陷的概率分割. 具体地, 将 Dropout 机制融入到改进的 Xception 网络, 使得从后验分布中生成蒙特卡罗样本; 其次, 提出

多尺度与多速率的空洞空间金字塔池化 (Atrous spatial pyramid pooling, ASPP) 模块, 提取任意分辨率下的密集特征图谱; 更简单有效的解码器细化目标的边界, 计算 Softmax 概率的均值和方差以作为分割预测和不确定性. 同时, 针对钢轨数据集中固有的类别不平衡问题, 提出在线前景-背景挖掘思想, 使用注意力网络 (LAN) 预测缺陷以获得加权系数, 从而补偿和抑制 Deeplab 权重训练的前景 Loss 分量和背景 Loss 分量. 注意力机制重新分配了主体网络的学习能力, 降低了原始任务的难度, 从而快速获得最优权重和精确的分割预测.

2 贝叶斯神经网络: DeeperLab

2.1 DeeperLab 近似推断

对于贝叶斯 DCNN, 感兴趣的是找到基于训练

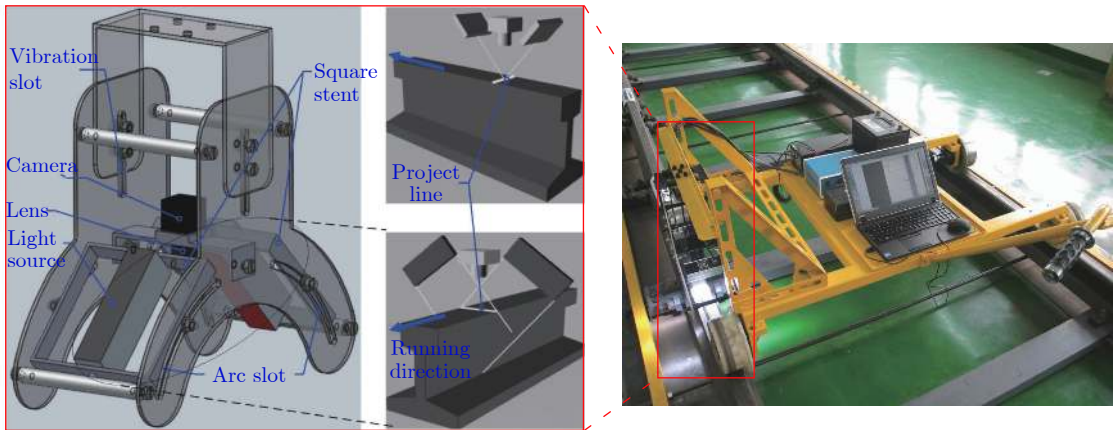


图 1 钢轨表面成像系统
Fig. 1 Rail surface imaging system

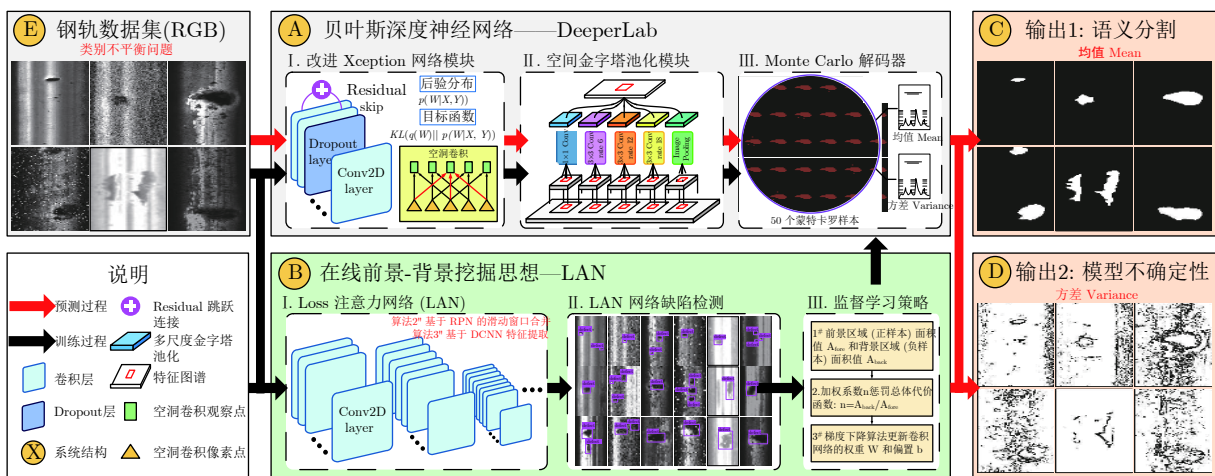


图 2 本文提出缺陷检测算法的整体框架
Fig. 2 Overview of the proposed rail defect detection algorithm

样本 X 和对应标签 Y 下, 卷积权重 W 的后验分布

$$p(W|X, Y) \quad (1)$$

然而, 这个后验分布是不可处理的, 因此使用变分推断来估计网络权重的分布. 这里, $q(W)$ 表示近似后验的变分分布. Gal 等^[19] 观察到, 通过网络权重定义的变分分布是一种伯努利分布, 因此在每个 $K \times K$ 维的卷积层 i 单元节点 j 上, $q(W_i)$ 定义为

$$\begin{aligned} b_{i,j} &\sim \text{Bernoulli}(p_i), \quad j = 1, 2, \dots, K \\ W_i &= M_i \text{diag}(b_i) \end{aligned} \quad (2)$$

其中, 利用伯努利分布随机变量的 b_i 向量和变分参数 M_i , $q(W)$ 被证明是文献 [19] 中高斯过程的近似模型. 有了真实后验和变分分布后, 定义目标函数为 K-L 散度项

$$KL(q(W)||p(W|X, Y)) \quad (3)$$

在文献 [19] 中, 作者还指出将式 (2) 中 p_i 的伯努利分布设置在隐藏层的权重上时, 相当于该层以 p_i 的 Dropout 率进行丢失. 因此, 本文提出一个新的轻量级、可伸缩性的 Dropout 网络, 不仅有效避免过拟合, 而且使用随机梯度下降而最小化 K-L 目标函数, 从而更好地学习权重数据的后验分布, 没有额外的参数化.

与通常的做法不同, 即使在测试阶段, 这些 Dropout 层也保持活力. 该思想是为了从后验分布中获取样本, 当 Dropout 层在网络权重上放置伯努利分布时, 训练网络上的随机前向遍历可以解释为从后验分布中生成蒙特卡罗样本. 因此, 同一输入的多个前向传递将产生很多 Softmax 类别概率, 然后其均值可以作为分割预测, 方差可以解释为每一类的模型不确定性.

2.2 编码器 - 解码器架构

提出 DeeperLab 的骨架是类似于 DeepLab v3+^[25]. 首先, 利用一种嵌有 Dropout 的空洞卷积或扩张卷积^[26] 的网络, 在不增加参数和不适用池化层的情况下, 扩大了输入特征图的视野和保持原有分辨率. 其次, 使用空洞空间金字塔池化 (ASPP)^[27] 和编码器 - 解码器的方法解决图像中存在的不同尺度对象问题. 虽然网络中没有池化层, 但是多个 stride 为 1 或者更高的卷积层降低了输出维度. 为了恢复原始尺寸, 可通过完全连接的条件随机场 (Conditional random field, CRF)^[28]、双线性插值^[29] 或可学习参数的解码器^[30] 的方法调整预测分割和不确定性的尺寸.

具体地, 如图 3 所示, 输入图像首先通过一个

改进的 Xception 网络^[31], 和多尺度图像处理的 ASPP 模块, 最后使用解码器调整图像到原始输入维度和产生尖锐的分割结果, 并将 Softmax 类别概率的方差输出为模型的不确定性. 本文提出的 DeeperLab 架构与文献 [25] 中提出的原始 DeeperLab v3+ 及其以往版本的区别如下:

1) Xception 的不同网络层存在 Dropout. 特别在中间流, 每 4 个模块单元后插入 1 个 Dropout 率为 0.5 的 Dropout 层. 中间流共有 16 个单元模块即 DeeperLab 共有 4 个 Dropout 层. 此外, 测试时 Dropout 用于蒙特卡罗采样, 生成类别标签的后验分布, 没有额外的参数化.

2) 本文网络中, 不使用级联的空洞模块或图像金字塔, 而是采用多尺度 ($\text{Conv2D} = 1 \times 1, 3 \times 3$) 和多速率 ($\text{Rate} = 6, 12, 18$) 的空洞空间金字塔池化 (ASPP) 以提取任意分辨率下的密集特征图谱 ($\text{Output_Stride} = 16$), 这种方法包含了更丰富的上下文语义信息和减少训练时间.

3) 一个更简单有效的解码器获得概率分割和模型不确定性. 与直接双线性 16 倍上采样不同, 首先对编码器特征进行 4 倍的双线性上采样, 然后与具有相同空间分辨率的低维特征连接 (如 Xception 的 Entry flow), 最后使用 3×3 卷积细化特征和通过另一个 4 倍的双线性上采样.

2.3 改进的 Xception 网络

本文中, 提出改进的 Xception 作为基础网络来实现 DeeperLab, 不仅使用了类似 VGG-16 的多层卷积叠加和类似 ResNet-101 的跳跃连接, 而且假设了在空间维度上 (高度和宽度) 执行的卷积操作与在深度维度上 (跨通道) 执行的卷积操作是分离的. 此外, 将 4 个 Dropout 层插入网络的中间流实现测试阶段的蒙特卡罗采样, 从而产生类别标签的后验分布.

如图 4 所示, 本文中新的轻量级、可伸缩性的改进 Xception 网络与原始的 Xception^[31] 的区别:

1) 没有池化层. 所有池化层由步长为 2 的深度卷积替换, 有利于密集预测. 根据 FCN^[32], 改进 Xception 网络是完全卷积的, 因此可以分割任意大小的图像.

2) 中间流含有更丰富的单元模块和 Dropout 层. 每 4 个单元模块后插入 Dropout 率为 0.5 的 Dropout 层, 而且共有 16 个模块并不是原始的 8 个模块. 额外的批量归一化和 ReLU 激活被增加到每个深度卷积后, 类似于 MobileNet.

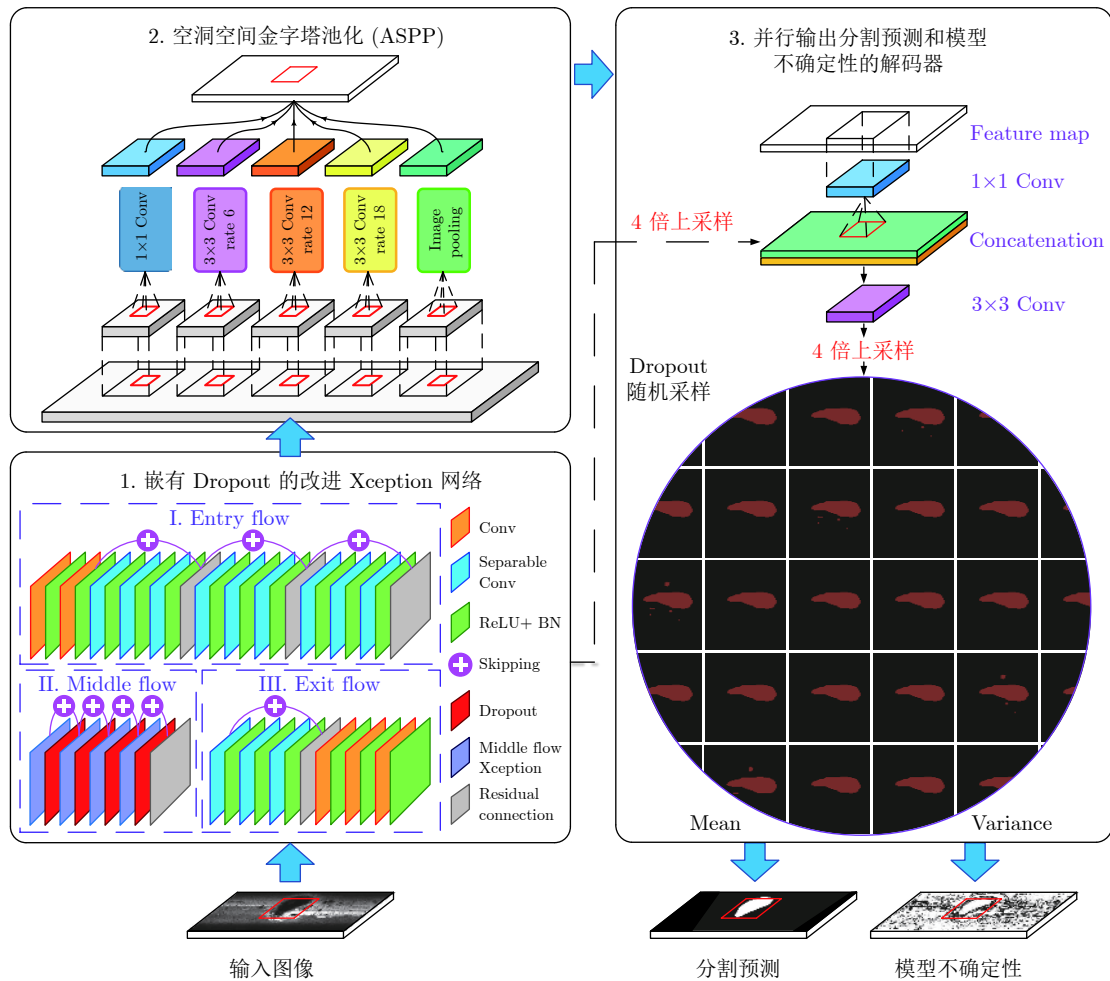


图 3 贝叶斯网络 DeeperLab 的编码器 - 解码器架构

Fig. 3 Encoder-decoder architecture of the proposed DeeperLab

3 注意力机制: LAN 网络

3.1 动机

3.1.1 问题陈述

正常情况下, 正负样本的比例应是较为接近的, 现有的分类模型、分割模型正是基于这一假设. 然而, 在特殊场景或特定任务下, 正负样本的比例却是相差悬殊, 存在严重的类别不平衡问题, 例如医疗肿瘤诊断和工业裂缝检测.

在本研究中, 使用上述 DeeperLab 进行钢轨表面缺陷分割是一项典型的二分类任务. 经验上, 缺陷前景的区域 (即负样本) 是远小于以钢轨背景的区域 (即正样本), 而且多尺度和多规格的表面缺陷是不可预测的. 因此, 钢轨数据的前景 - 背景不平衡问题使得模型在训练中无法充分学习到缺陷附近的上下文信息, 虽然 DeeperLab 的 ASPP 缓解了这个难题. 其次, 简单易分的背景太多, 极有可能掩盖

其他类别 (即缺陷前景) 的作用, 详细的解释如下.

首先, DeeperLab 的损失函数由前景类别和背景类别的 Loss 值组成, 即

$$L_{\text{total}} = L_{\text{fore}} + L_{\text{back}} \quad (4)$$

虽然前景仍产生一定幅度的 L_{fore} , 但是背景产生的 L_{back} 将对总 Loss 值是主导效果, 从而主导梯度的更新方向和掩盖重要信息, 导致训练中总体 L_{total} 下降缓慢. 因此, 使用权重 W 补偿和抑制不同类别的 Loss 值是关键, 现在更新加权的损失

$$L_{\text{total}} = W_{\text{fore}} \times L_{\text{fore}} + W_{\text{back}} \times L_{\text{back}} \quad (5)$$

3.1.2 解决方案

本文提出一种在线前景 - 背景挖掘 (Online foreground background mining, OFBM) 的思想.

具体地, 当每批量的数据序列输入时, 损失注意力网络 (LAN) 首先定位缺陷目标的边框, 然后获得前景与背景的加权系数并传递给执行分割任务的 DeeperLab, 最后通过加权 Loss 的方式补偿前

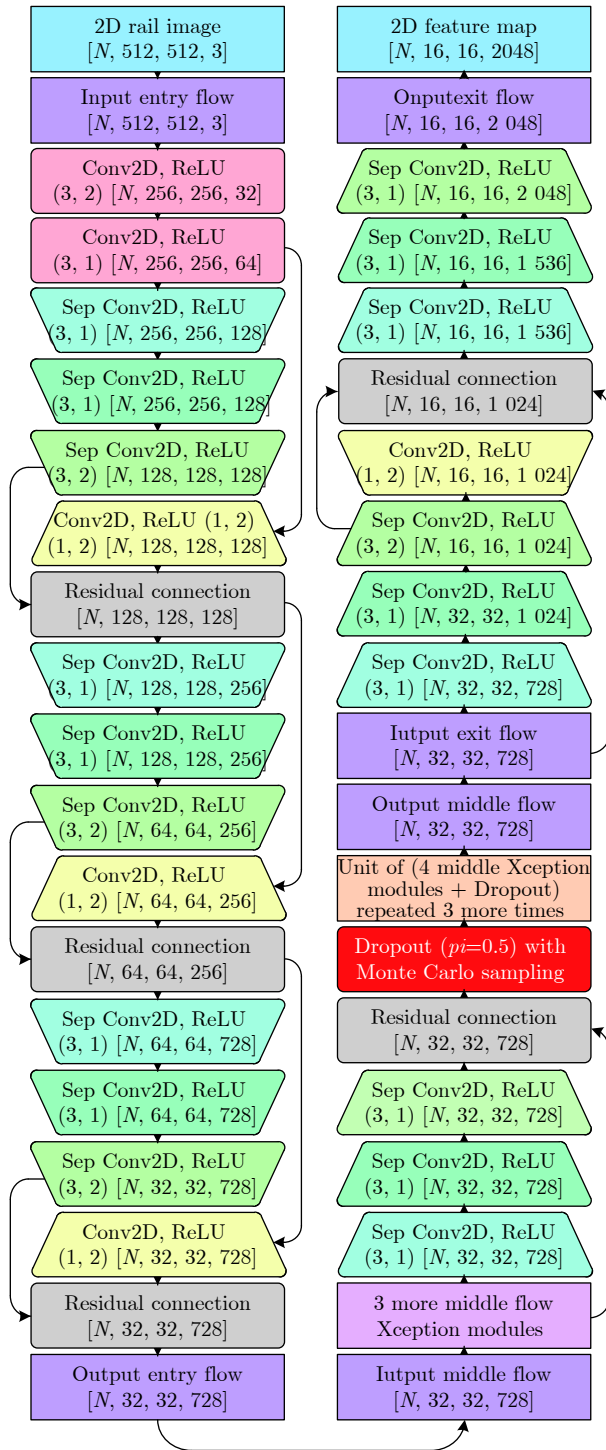


图 4 嵌有 Dropout 的改进 Xception 网络

Fig. 4 Improved Xception network with Dropout

景, 同时抑制背景. 这种注意力机制重新分配了网络的学习能力, 降低了原始任务的难度, 使网络更加容易训练和 Loss 值下降更快. 值得注意. 算法 1 仅阐述了单批次下的参数学习过程.

算法 1. 基于注意力机制的 DeeperLab 权重学习

输入. 训练数据 $X = (N_x, W_x, H_x, 3)$, 标签 Y

输出. 卷积网络权重 $W_{network}$

步骤 1. 损失注意力 LAN 定位目标的边界框 $B = (X_b, Y_b, W_b, H_b)$;

步骤 2. 计算前景类别的面积大小 A_{fore} : $A_{fore} = W_b \times H_b$;

步骤 3. 计算背景类别的面积大小 A_{back} : $A_{back} = W_x \times H_x - A_{fore}$;

步骤 4. 获得前景和背景的加权系数 n : $n = W_{fore} / W_{back} = A_{back} / A_{fore}$;

步骤 5. 在式 (5) 中分别补偿和抑制 DeeperLab 的前景 Loss 分量和背景 Loss 分量;

步骤 6. 使用梯度下降算法更新卷积权重 $W_{network}$. 算法结束.

3.2 注意力网络 LAN

图 3 是损失注意力网络 (LAN) 的架构, 本质上它是基于 YOLO (You only look once)^[33] 的目标检测深度学习. 根据输入的图像数据, 运用回归方法输出多个滑动窗口的位置和该窗口内检测到的目标类别.

首先, 将输入图像划分为 $S \times S$ 个单元格, 区别于其他模型, 这里扩展了网格尺寸 $S = 14$ 以增强对小目标特征的提取能力, 实现多尺度目标识别, 从而提高缺陷检测的准确性. 在实践中, 每个单元格的神经元负责检测“落入”该单元格的对象, 通常涉及到多个滑动窗口. 为减少空间冗余和时间成本, 提出了基于区域提议网络 (Region proposal network, RPN) 的滑动窗口合并算法, 以确定滑动窗口的边框. 具体参见算法 2.

算法 2. 基于 RPN 的滑动窗口合并算法

输入. 图像数据 X_{pic}

输出. 检测目标的滑动窗口集合 L

步骤 1. 使用网格划分的方法将 X_{pic} 划分为 n 个单元, 生成集合 $R = \{S_1, S_2, \dots, S_n\}$;

步骤 2. 初始化单元格 S_i 的相似集合 $m_i = 0$, 并初始化 14×14 规格的滑动窗口;

步骤 3. for 滑动窗口中的邻近区域对 (S_i, S_j) do

a) 采用 RPN 方法及式 (1) 计算滑动窗口内与 S_i 相邻的所有单元格 S_j 的特征相似度 $F(S_i, S_j)$;

b) 找出最大相似度 $F_{max}(S_i, S_j)$;

c) 更新单元格 S_i 的相似集合 $m_i = m_i \cup \{F_{max}(S_i, S_j)\}$;

d) while 每个单元格 S_i 的相似集合 $m_i \neq 0$

i) 找出集合 m_i 中的元素对应的所有单元格, 并去除不包括检测对象的单元格;

ii) 将所获得的单元格与单元格 S_i 合并成

新的 S_i , 并将其作为集合 L 的元素.

步骤 4. 输出目标位置检测滑动窗口集合 i . 算法结束.

若 B 为每个单元格可以用于预测对象的滑动窗口数量, C 为类别总数, 则 LAN 的全连接层的输出维度是: $S \times S \times (5 \times B + C)$.

LAN 的损失函数计算式为

$$\lambda_{\text{loss}} = \sum_{i=1}^{S^2} E_c(i) + E_{\text{IOU}} + E_{\text{class}} \quad (6)$$

其中, 图像被划分为 $S \times S$ 个网格, $E_c(i)$ 表示第 i 个网格产生的类别误差, E_{IOU} 和 E_{class} 分别为预测数据和标定数据之间的坐标误差.

图 5 包含 18 个特征提取的卷积层、5 个缩减图像像素的池化层、1 个预测位置和估计类别的全连接层和 Softmax 输出, 而且 BN 和 ReLU 均设置在每个卷积层后, 全连接层采用 Dropout 随机丢弃神经元. 每个单元格涉及多个滑动窗口 $T(x, y, w, h, c)$, 其中 x 和 y 分别是当前网格检测对象中心位置的横坐标和纵坐标, w 和 h 分别是滑动窗口的宽度和高度, c 是置信度. 算法 3 给出基于 YOLO 的特征提取算法.

算法 3. 基于 YOLO 的特征提取算法

输入. 图像数据集 X

输出. 图像数据 X 的特征模型 M_{weights}

步骤 1. 数据预处理. 针对图像数据集 X 的每一幅图像采用 LabelImg 工具获得真实目标的矩形区域坐标, 生成每幅图像中真实目标的坐标信息文件 F_c ;

步骤 2. 加载 LAN 的图像分类训练模型, 同时初始化图片数据 X 的特征模型 M_{weights} , 初始化每幅图像的预测矩形区域坐标为空;

步骤 3. 坐标信息文件 F_c , 基于 RPN 方法生成每幅图像的若干个目标候选区域矩阵向量;

步骤 4. 将候选区域矩阵向量作为第 1 层的输入, 将其结果作为第 2 层的输入;

步骤 5. 执行池化操作;

步骤 6. 将步骤 5 中的结果作为输入, 采用一个滑动窗口扫描网格, 进行卷积与池化操作计算出滑动窗口内单元格的特征向量;

步骤 7. 将步骤 6 所得的特征向量作为第 18 个卷积层的输入, 运用 2×2 步长进行卷积操作;

步骤 8. 将步骤 7 的输出作为全连接层的输入, 采用 1×1 步长进行卷积操作;

步骤 9. 将步骤 8 的输出作为分类函数 Softmax 的输入, 计算图片数据 X_{pic} 的预测概率估计值 P_{pic} , 并保存运用算法 2 获得的重叠面积最大的 P_{IOU} 对应的目标区域特征;

步骤 10. 将对应的目标区域的特征保存到特征模型 M_{weights} 中每一个类别相对应的位置;

步骤 11. 输出特征模型 M_{weights} . 算法结束.

4 实验与分析

本文实验中, 深度网络是在 Ubuntu 16.04, GTX Tian XP GPU, Intel Xeon 24 core CPU, 64 GB 和 TensorFlow 框架的环境下训练和预测, 指标评估是在 MacBook Pro, 3.1 GHz, Intel core i5, 8 GB 的 MATLAB 2017a 环境下数值运算. 原始数据集包含 4 000 张 RGB 钢轨样本, 其中数据增强方案模拟了实际工程中的表面场景和外界环境, 即在实验室条件下通过 HSV 指数因子的对比度变换、多角度的旋转变换、缩放变换、噪声扰动、翻转变换、尺度变换以模拟不同光源、成像焦距、随机噪声分布、设备振动的合成图像, 从而使得原始数据集扩充至 11 倍(即 44 000 张样本), 其中训练集、验证集和测试集的比例为 7 : 2 : 1. 此外, 批量尺寸为 16, DeeperLab 训练的学习率为 0.001, Xcep-

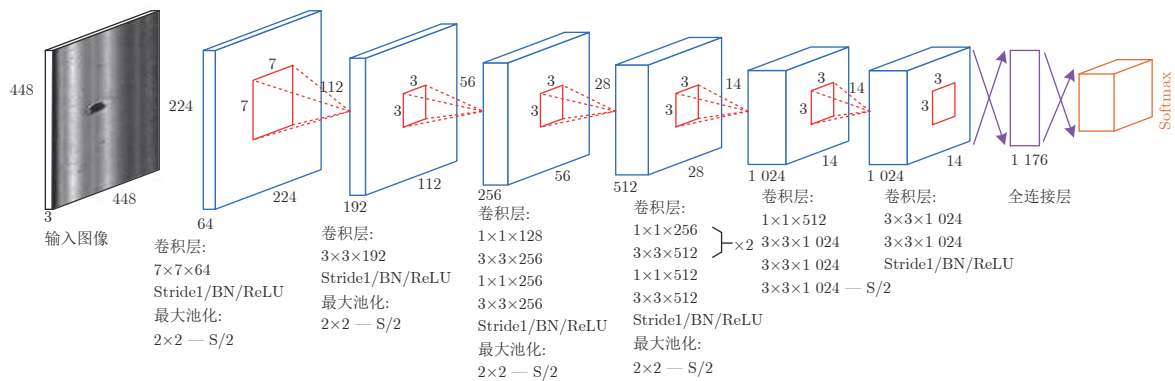


图 5 损失注意力网络 (LAN) 的结构示意图

Fig. 5 Structure of the proposed loss attention network (LAN)

tion 网络的 Dropout 率为 0.50, 以防止过拟合和从后验分布中产生蒙特卡罗样本, LAN 训练的学习率为 0.01, 全连接层 Dropout 率为 0.60, 采用 Early Stopping 策略以避免特征过拟合。

4.1 LAN 性能测试

算法 1 中, 提出 LAN 以指引 DeeperLab 如何补偿和抑制前景 Loss 和背景 Loss, 从而解决类别不平衡的问题。因此, 稳定和精确的 LAN 网络非常重要。

4.1.1 基于单样本图像的评估

图 6 给出了基于钢轨条件的 LAN 网络检测案例。这里研究表面缺陷的缺陷类型, 其中由面积大小可分为轻微、中等和重度缺陷。钢轨场景 I~VI 是指洁净光滑的表面、人为高斯噪声的光滑表面、粗糙生锈的表面、人为椒盐噪声的生锈表面、过度曝光表面和光照不足表面。GT 标记在左原始图像, LAN 检测结果标记在右预测图像。对于由好至差的结果, 评估等级分别是完全检出 (Fully detected)、潜在误差 (Latent error) 和丢失 (Missing)。

分析可知: 1) 大体上, 在各类钢轨场景中 LAN 网络有效稳定地检测不同尺度的表面缺陷, 因为有大量完全检出的结果输出; 2) 14×14 规格的扩增网格增强了 LAN 网络对小目标的特征提取能力, 实现相邻很近或者尺度很小的多目标识别, 如第 1 行 I 和 II。但是, III 和 VI 表明它对铁锈干扰和过暗光照的条件非常敏感; 3) 基于 RPN 的滑动窗口合并算法提高了 LAN 对中等与重度缺陷的检测能力。特别是在第 2 行 III、IV 和第 3 行 I、III、IV 的复杂背景中, 多比例和多尺度的表面缺陷均被精确识别。

应用 LAN 网络在整个测试数据集, 包括塌陷、裂纹、剥离、断轨类型的表面缺陷。图 7 定量给出了

基于钢轨条件的 LAN 箱形图。分析可知: 1) 大体上, 在各类钢轨场景中 LAN 有效稳定地检测不同尺度的表面缺陷, 因为有大量高于 95% 的中位线和非常窄的矩形输出; 2) 14×14 规格的扩增网格增强了 LAN 网络对小目标的特征提取能力, 以接近 95% 的中位线。然而, 轻微缺陷 (图 7(a)) 里场景 III、IV 的不可预测的 5 个低于 60% 的异常值表明它对具有挑战性的生锈表面非常敏感, 而且 VI 的矩形比其他的都长; 3) LAN 对中等和重度缺陷的检测效果最好。虽然 (图 7(b)) 场景 V 的过度曝光条件导致相对低的中位线和相对长的矩形, (图 7(c)) 场景 II 的高斯噪声条件导致很低的异常值, 但是平均高于 90% 的中位线验证了 LAN 准确性和稳定性。

4.2 DeeperLab 性能测试

训练时, LAN 辅助监督 DeeperLab 权重学习。测试时, 活跃的 Dropout 前向遍历, 生成蒙特卡罗样本。最后输出均值为分割预测, 方差则解释为每一类的模型不确定性。这是非贝叶斯 CNN 无法做到的。

为了测试提出的算法, 下面提供了多组对比实验与结果分析。不同模型 (FCN, Unet, SegNet, PSPNet, DeepLab v3+, Mask RCNN 和本文方法) 的超参数实现优化, 其中网络训练的学习率依次为 0.01, 0.001, 0.03, 0.01, 0.0001, 0.005 和 0.001, 批量大小依次为 64, 64, 32, 32, 32, 32 和 16, 此外, Momentum 动量设为 0.98, 最大迭代次数为 30 000 步, 使用 Relu 激活函数避免梯度消失。

4.2.1 基于单样本图像的评估

图 8 是本文提出方法和其他方法的测试对比结

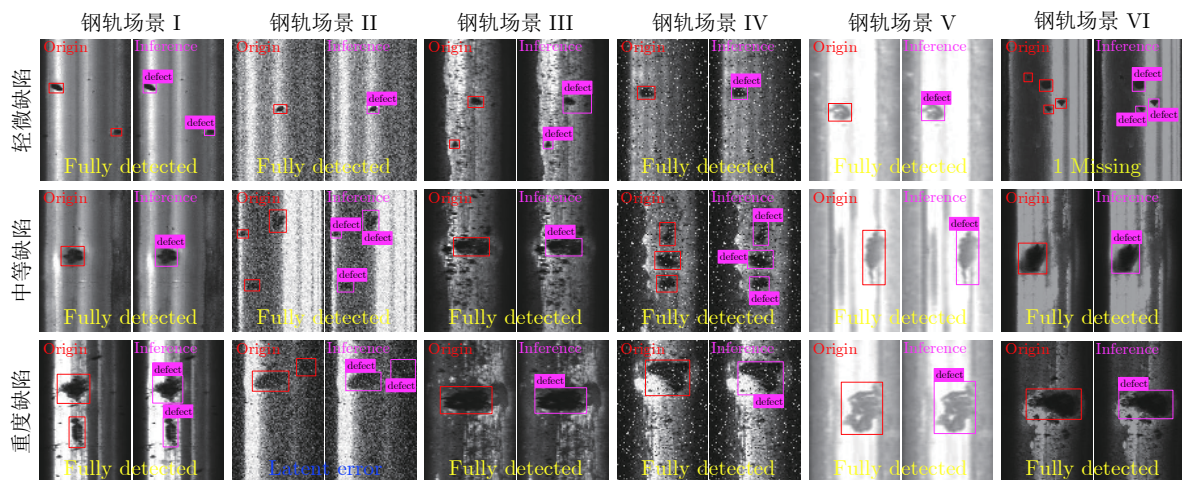


图 6 不同钢轨场景下 LAN 网络对不同尺度表面缺陷的检测结果

Fig. 6 LAN detection results of different scaled defects in various rail scenes

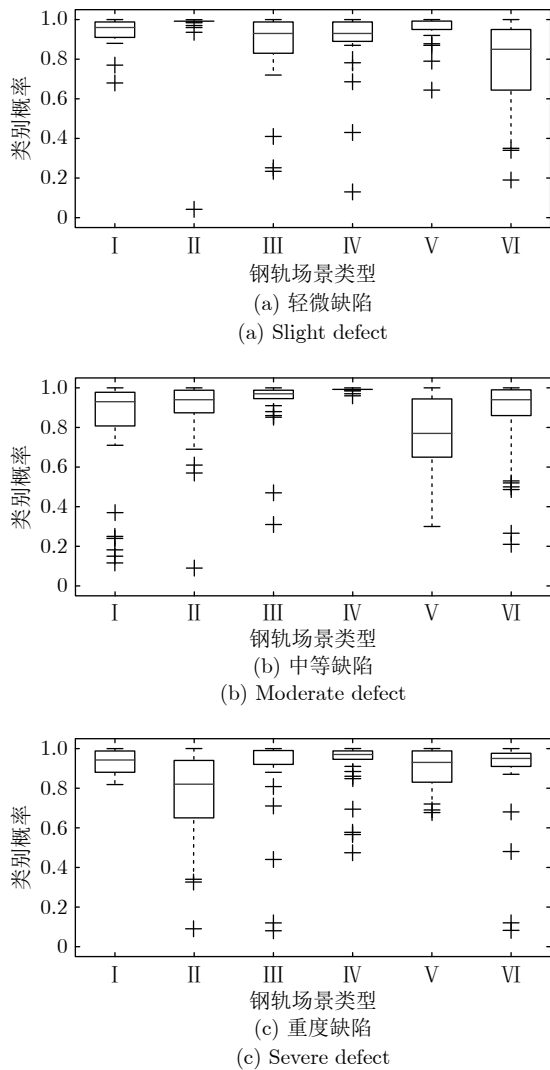


图 7 不同条件的 LAN 测试箱形图

Fig. 7 Box-plot of LAN in different conditions

果. 其中 (a)~(f) 为场景类型 I~VI 的钢轨样本图, (1)~(4) 为输入图像、GT、本文方法分割结果和可视化的不确定性估计, 其中蒙特卡罗采样次数为 50 次, 方差阈值为 0.3, (5)~(11) 为 FCN^[32]、Unet^[34]、SegNet^[35]、PSPNet^[36]、之前工作^[29]、DeepLab v3+^[25]和 Mask RCNN^[23] 的分割结果, 其中 FCN 使用 32 倍上采样, Mask RCNN 忽略界框. 从中可知: 1) 本文方法具有最好的多尺度缺陷的分割精度和多类型场景的鲁棒性. 这是因为新增 Dropout 的随机前向遍历产生了蒙特卡罗样本, 使得解码器充分学习网络权重的后验分布和执行近似贝叶斯推理, 从而更好地细化缺陷的边缘细节. 2) 虽然没有定量指标来评价不确定性的好坏, 但是子图 (a4)~(f4) 定性地可视化了模型对于其预测的置信度. 其中, 特别在缺陷边缘、表面铁锈油污和加性噪声等区域的预测没有很好的信心. 3) 本文钢轨数据

集的应用中, FCN 具有更加线性粗糙的目标边界, Unet 出现错误识别和模糊边界, SegNet 非线性识别的对象尺度比 GT 更大, PSPNet 准确性较好但出现了毛刺和噪点, 之前工作出现了严重的错误类别和噪声, DeepLab v3+ 因为无法解决数据不平衡而导致识别率很低, 而 Mask RCNN 的效果总体上是相对最好.

表 1 记录了定量的测试结果, 其中误分类率 (Miss classified ratio, MCR)、边缘索引 (Rand index, RI)、峰值信噪比 (Peak signal to noise ratio, PSNR)、Jaccard 指标 (Jacc) 和信息变分 (Variation of information, VI) 用于评价, 最好和次好的性能数据分别由粗体和斜体字体标出. 可以观察到, 本文方法具有很高的分割精度和鲁棒性优势.

4.2.2 基于测试数据集的评估

应用本文方法和其他对比方法在整个钢轨数据集, 包括洁净光滑的表面 (场景 I)、高斯噪声的光滑表面 (场景 II)、粗糙生锈的表面 (场景 III)、椒盐噪声的生锈表面 (场景 IV)、过度曝光 (场景 V) 和光照不足的表面 (场景 VI). 缺陷类型包括塌陷、裂纹、剥离和断轨. 利用 MATLAB 2017a 曲线拟合工具箱绘制 P-R 曲线, 如图 9 所示. 从中观察到: 1) 本文方法鲁棒有效地检测所有钢轨场景的表面缺陷. 例如在钢轨表面 III, 若有 80% 缺陷被检出 (查全率 = 80%), 本文方法的精度 (查准率 = 90%) 是 PSPNet 的 3 倍 (查准率 = 70%) 和 SegNet 的 5 倍 (查准率 = 50%); 2) 因为简单场景 I 和 II 与缺陷尺寸相对统一, 之前工作、PSPNet、DeepLab v3+ 和 MaskRCNN 都表现很好的分割性能, 但是对于恶劣条件下的场景 III~VI, 它们受到多尺寸缺陷、加性噪声、铁锈油污、曝光或光源不足的严重干扰而失效; 3) FCN 分割非常线性粗糙的目标边界, Unet 获得更加模糊的缺陷边缘和错误识别对象, SegNet 非线性识别的对象往往比 GT 更大, 导致它们的检测性能最差 (查准率 = 40%~60%).

4.3 不同消融研究

4.3.1 概率变体的消融研究

理想情况下, 在 DeeperLab 的每个隐含层后都可以训练一个完整的贝叶斯网络. 然而, 这是一个过于强大的正则化, 导致网络学习非常缓慢. 因此, 研究不同编码器和解码器 Dropout 配置的不同贝叶斯变体. 具体如下:

1) 贝叶斯编码器: 将 Dropout 插入每个编码器单元.

2) 贝叶斯解码器: 将 Dropout 插入每个解码器单元.

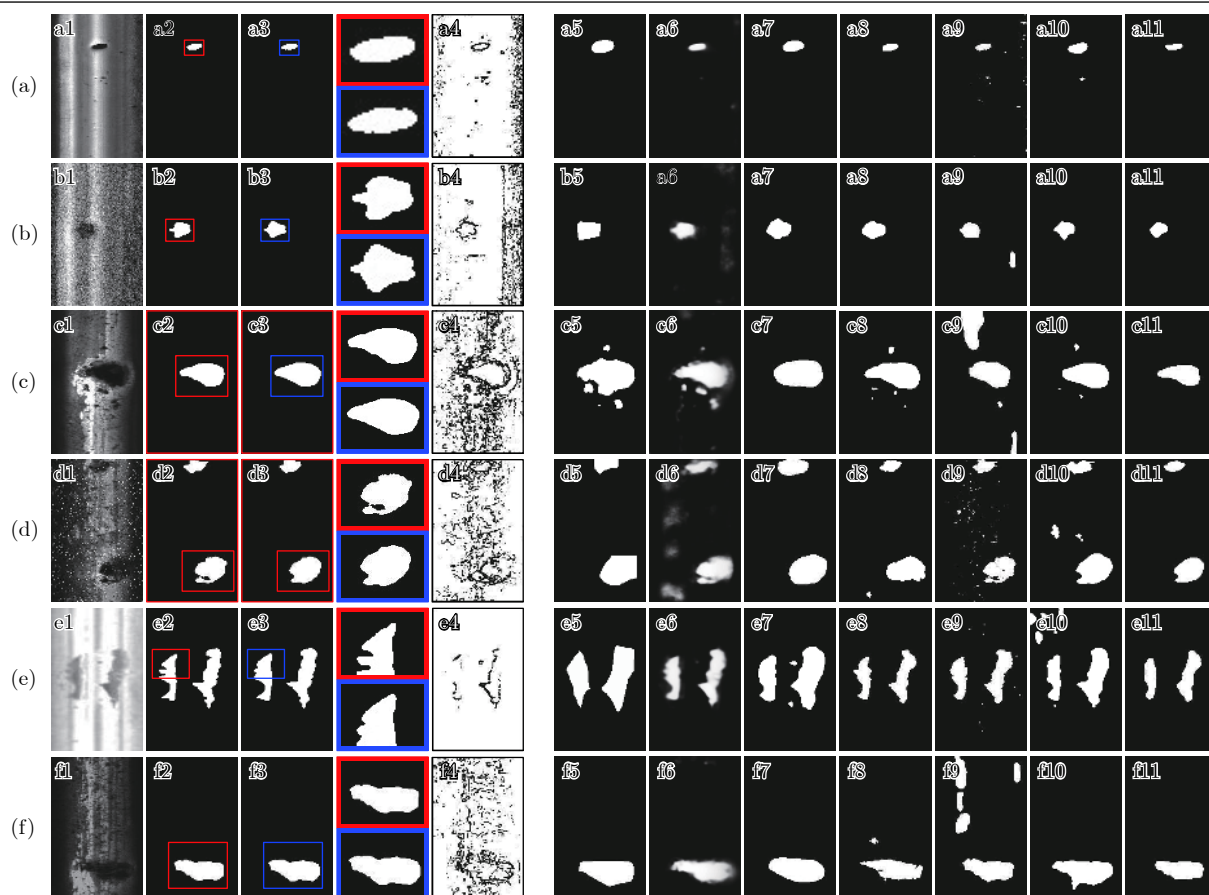


图 8 本文方法和其他方法在不同钢轨样本的测试结果

Fig. 8 Results of the proposed method and other methods on various rail samples

3) 贝叶斯编码器 - 解码器: 将 Dropout 层插入到每个编码器和解码器单元.

4) 贝叶斯输入流: 在 Xception 网络的输入流, 每 1 个输入流模块后插入 1 个 Dropout 层. 由于输入流有 3 个模块, 因此这个变体总共有 3 个 Dropout 层.

5) 贝叶斯中间流: 在 Xception 网络的中间流, 每 4 个中间流模块后插入 1 个 Dropout 层. 由于中间流有 16 个模块, 因此这个变体总共有 4 个 Dropout 层.

6) 贝叶斯输出流: 在 Xception 网络的输出流, 每 1 个输出流模块后插入 1 个 Dropout 层. 由于输出流有 2 个模块, 因此这个变体总共有 2 个 Dropout 层.

应用钢轨数据集在这些贝叶斯变体, 比较 50 个蒙特卡罗样本的加权平均性能. 可定性观察到所有变体都输出相似的模型不确定性, 说明它们在分割的边界附近是不确定的, 而且具有视觉上不明确的对象, 例如 离散细微的铁锈和连通成块的阴影. 此外, 表 2 给出了定量的分割性能差异, 其中, Jacc (%) 为 Jaccard 指标, Dice (%) 为 Dice 指标, 加权平均是指在测试时去掉 Dropout 层, 并将网络

权重按 Dropout 率缩放.

从中可见: 1) Dropout 所有编码器或解码器都是过度正则化, 导致很低的训练拟合和很差的分割性能; 2) 贝叶斯输入流的测试结果一般, 因为浅层网络的低水平特征 (如缺陷的边和角) 在模型分布中是一致的, 因此它使用确定的权重表示; 3) 贝叶斯中间流的测试结果最好, 因为深层的高级特征 (如形状和上下文关系) 可以使用概率权重有效地建模; 4) 加权平均法得到更差的分割效果, 而且无法提高模型不确定性的度量. 虽然采样以运行时间作为代价, 但高性能 GPU 可降低此成本.

4.3.2 批量尺寸的消融研究

通常情况下, 在合理的范围内, 越大的批量尺寸 (batchsize) 越有助于网络充分学习, 使得 Loss 曲线下降更快和振荡越小, 最终获得最佳的网络权重. 然而在本文, 辅助的注意力网络 LAN 被引入对主体的贝叶斯深度模型 DeeperLab 进行监督优化, 单个周期内不同规格的数据流将决定以何种程度拟合前景 - 背景的加权系数, 因此批量尺寸决定了训练后的网络为无注意力网络、欠注意力网络或过注

表 1 本文方法和其他方法在不同钢轨样本的定量结果
Table 1 Quantitative results of our method and other methods in various rail samples

图像	指标/方法	FCN ^[32]	Unet ^[34]	SegNet ^[35]	PSPNet ^[36]	之前工作 ^[29]	DeepLab v3+ ^[25]	Mask RCNN ^[23]	本文方法
样本 1	MCR (%)	2.25	11.28	1.87	1.12	4.71	1.33	0.94	1.01
	RI (%)	97.65	80.87	98.33	98.57	92.46	98.39	98.65	99.60
	PSNR (dB)	19.18	28.93	21.09	25.56	24.33	21.46	29.07	32.78
	Jacc (%)	31.86	41.67	41.92	64.06	58.35	44.15	81.55	91.09
	VI (pixel)	0.20	0.91	0.16	0.12	0.50	0.14	0.11	0.08
样本 2	MCR (%)	1.69	29.72	1.58	1.35	3.43	2.40	2.27	2.49
	RI (%)	98.19	53.90	98.31	98.65	96.63	98.42	98.36	98.61
	PSNR (dB)	19.37	23.50	19.97	21.49	20.01	24.76	24.53	26.07
	Jacc (%)	57.60	27.43	61.19	69.01	59.23	81.84	78.72	85.99
	VI (pixel)	0.16	2.24	0.17	0.13	0.31	0.17	0.17	0.15
样本 3	MCR (%)	6.79	31.77	8.95	3.24	14.78	7.91	6.73	6.85
	RI (%)	89.00	57.92	93.65	95.35	82.54	95.79	96.40	97.63
	PSNR (dB)	12.12	16.95	15.40	15.81	11.41	18.29	19.89	23.37
	Jacc (%)	46.47	38.81	64.76	66.97	39.73	78.69	81.00	91.99
	VI (pixel)	0.52	2.65	0.44	0.31	1.04	0.35	0.35	0.24
样本 4	MCR (%)	4.66	45.13	10.85	4.45	14.95	11.25	8.43	8.51
	RI (%)	94.41	44.64	92.49	13.62	84.56	91.29	95.16	96.10
	PSNR (dB)	14.93	16.94	14.15	15.98	19.97	14.96	19.80	21.53
	Jacc (%)	64.13	31.80	60.45	67.33	83.72	64.04	82.88	88.98
	VI (pixel)	0.44	3.67	0.61	0.47	1.16	0.68	0.49	0.42
样本 5	MCR (%)	7.83	23.93	7.61	12.30	13.98	14.07	11.98	12.21
	RI (%)	89.42	76.97	89.73	92.63	91.11	89.75	91.80	95.91
	PSNR (dB)	12.34	19.27	12.53	16.67	16.36	13.42	15.80	19.98
	Jacc (%)	59.42	70.08	61.14	76.56	76.15	65.02	71.83	89.84
	VI (pixel)	0.68	2.09	0.66	0.81	1.00	0.66	0.79	0.51
样本 6	MCR (%)	9.00	17.68	8.64	8.31	14.30	9.54	6.60	7.35
	RI (%)	94.33	79.14	95.12	94.87	84.60	93.16	98.00	97.44
	PSNR (dB)	16.11	20.54	16.56	17.70	12.45	15.80	22.38	22.64
	Jacc (%)	69.14	59.89	71.84	74.68	49.52	67.57	89.78	91.15
	VI (pixel)	0.45	1.62	0.40	0.47	0.98	0.53	0.28	0.26

注意力网络. 图 10 中, (a1)~(a8) 是原始图像、GT、 $\beta = 2$ (MCR = 16.12%)、 $\beta = 4$ (MCR = 8.20%)、 $\beta = 8$ (MCR = 1.21%)、 $\beta = 16$ (MCR = 0.87%)、 $\beta = 32$ (MCR = 1.64%)、 $\beta = 64$ (MCR = 2.15%) 分割结果, (b1)~(b8) 是 (a1)~(a8) 中矩形框细节.

从中可知, 当批量很小时 ($\beta = 2$), LAN 欠拟合前景 - 背景的加权系数, 使得在每次迭代 DeeperLab 的 Loss 下降方向和步长的随机性很大, 出现振荡而难以达到收敛, 最终预测出错误类别和大量噪声, 说明当前权重是欠注意力网络. 当批量很大时 ($\beta = 64$), LAN 过拟合了前景 - 背景的加权系数, 使得 DeeperLab 的 Loss 值出现局部最优的情况而无法继续下降, 最终预测出过于平滑而没有边缘细节的

缺陷目标.

4.3.3 综合性能的消融研究

表 3 分析了在不同的特征提取的骨架、有/无数据扩增 (Data augmentation, DA)、有/无更大的蒙特卡罗采样规格 (Monte Carlo, MC) 的系统配置下, 所提出模型的综合性能. 从中观察到: 1) Xception65 骨架获得最好的识别精度, 因为它不仅具有多个层叠的卷积层和与 ResNet 类似的跳跃连接, 而且使用基于 Inception 的深度分离卷积; 2) 利用数据扩增和更大的蒙特卡罗采样规格有助于性能优化; 3) 虽然活跃的 Dropout 层在测试阶段进行的采样工作是以时间为代价, 例如在 250×160 样本图像上用时 180.53 ms, 但高性能 GPU 可降低此成本.

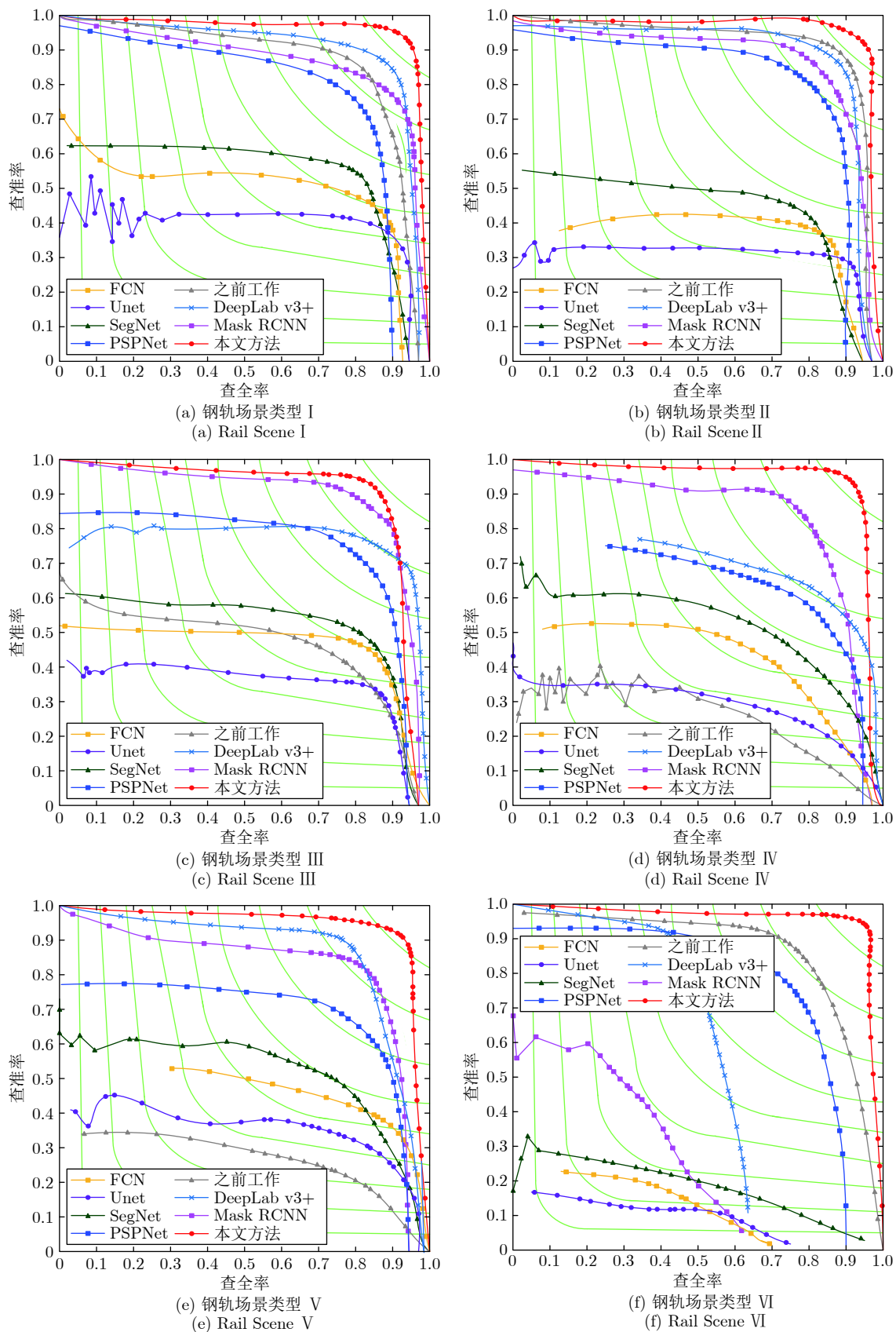


图 9 不同钢轨场景类型的 P-R 曲线
Fig. 9 P-R curves of different rail scene types

表 2 不同贝叶斯变体的性能 (%)

Table 2 Performance of different Bayesian variants (%)

概率变体	加权平均法		蒙特卡罗采样法	
	Jacc	Dice	Jacc	Dice
无 Dropout	68.36	68.95	—	—
编码器	55.24	56.71	64.60	66.07
解码器	61.78	61.34	63.92	65.88
编-解码器	58.62	60.12	60.57	62.49
输入流	75.44	76.21	82.65	80.33
中间流	83.12	80.69	90.43	91.52
输出流	68.50	67.33	77.21	78.06

5 结束语

针对复杂多样的钢轨场景, 本文扩展了最先进的深度学习框架 DeepLab v3+ 到一个新的轻量级、可伸缩性的贝叶斯版本 DeeperLab, 实现表面缺陷的概率分割. 具体地, 嵌有 Dropout 的改进 Xception 网络在预测时随机前向遍历, 使得从后验分布生成蒙特卡罗样本, 而没有额外的参数化. 此外, 空洞空间金字塔池化 (ASPP) 密集地提取全局和局部的上下文信息, 解码器输出 Softmax 概率的

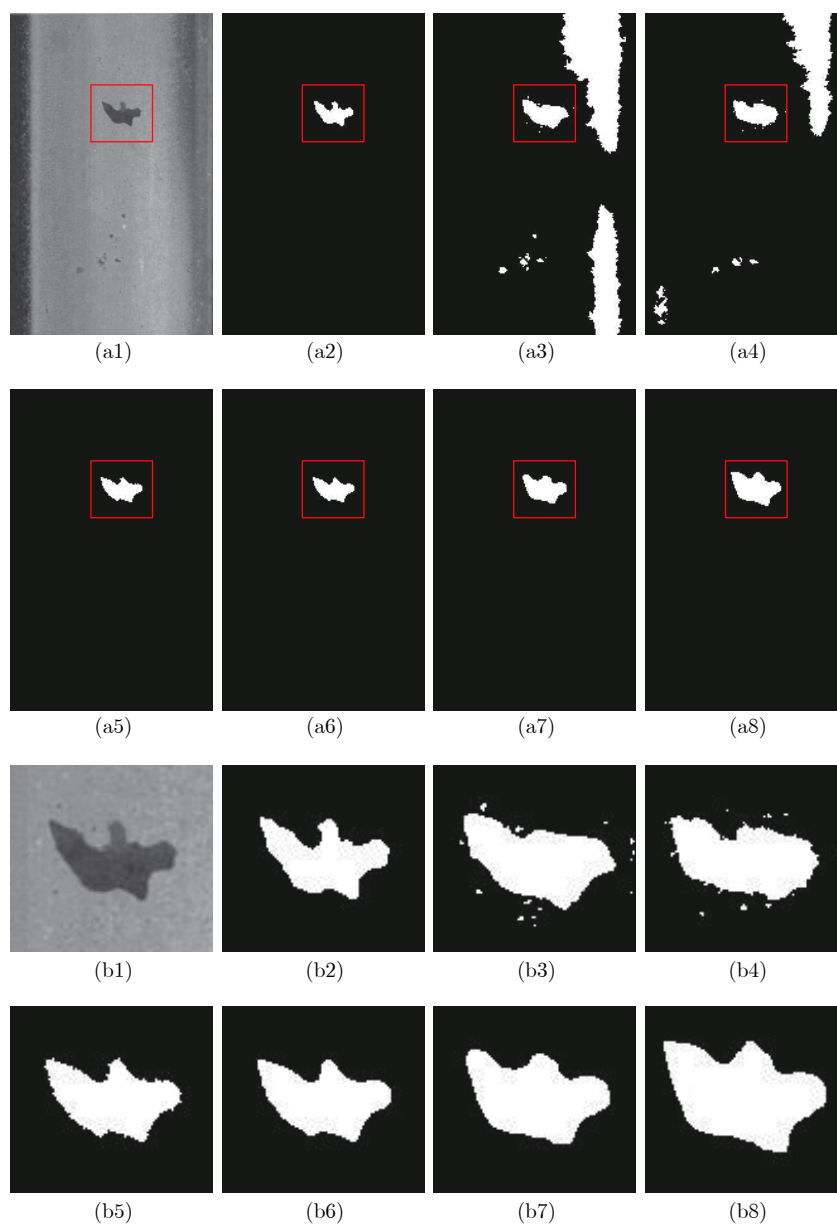


图 10 使用不同批量尺寸训练的测试结果

Fig. 10 Results of our method with different batch sizes

表 3 综合性能的消融研究
Table 3 Ablation experiment of comprehensive performance

方法	Pixel Jacc. (%)	运行时间 (ms)			模型成本(MB)	训练成本(GB)
		60 × 40	250 × 160	500 × 300		
MobileNet ($\beta = 16$)	77.17	19.91	53.10	133.49	23	3.82
ResNet50 ($\beta = 16$)	77.80	40.55	141.92	336.36	274	4.43
ResNet101 ($\beta = 16$)	78.45	66.37	181.80	431.42	477	6.99
Xception34 ($\beta = 16$)	81.66	46.64	149.13	352.70	288	3.97
Xception34 + DA ($\beta = 16$)	83.25	—	—	—	—	3.95
Xception65 + DA ($\beta = 16$)	88.73	79.64	159.29	517.70	439	4.20
Xception65 + DA + MC ($\beta = 16$)	91.46	90.26	180.53	586.73	—	5.56

均值和方差作为语义分割和模型不确定性, 这是其他 DCNN 无法做到的. 同时, 为解决类别不平衡问题, 提出在线前景 - 背景挖掘思想, 训练中利用注意力网络 LAN 获得加权系数, 从而补偿和抑制 DeeperLab 的前景与背景 Loss 分量, 实现辅助监督训练. 实验结果表明, 注意力机制重新分配了主体网络的学习能力, 降低了原始任务的难度, 从而快速获得最优权重和精确的分割预测. 本文方法具有 91.46% 分割精度和 0.18 s/帧的运行速度, 相比其他方法更加精确快速.

References

- 1 He Zhen-Dong, Wang Yao-Nan, Mao Jian-Xu, Yin Feng. Research on inverse P-M diffusion-based rail surface defect detection. *Acta Automatica Sinica*, 2014, **40**(8): 1667-1679 (贺振东, 王耀南, 毛建旭, 印峰. 基于反向 P-M 扩散的钢轨表面缺陷视觉检测. 自动化学报, 2014, **40**(8): 1667-1679)
- 2 He Z D, Wang Y N, Yin F, Liu J. Surface defect detection for high-speed rails using an inverse PM diffusion model. *Sensor Review*, 2016, **36**(1): 86-97
- 3 Resendiz E, Hart J M, Ahuja N. Automated visual inspection of railroad tracks. *IEEE Transactions on Intelligent Transportation Systems*, 2013, **14**(2): 751-760
- 4 Sun Ci-Suo, Zhang Yu-Hua. Research on automatic early warning method for rail flaw based on intelligent identification and periodic detection. *Journal of the China Railway Society*, 2018, **40**(11): 140-146 (孙次锁, 张玉华. 基于智能识别与周期检测的钢轨伤损自动预警方法研究. 铁道学报, 2018, **40**(11): 140-146)
- 5 Liang B, Iwnicki S, Ball A, Young A E. Adaptive noise canceling and time-frequency techniques for rail surface defect detection. *Mechanical Systems and Signal Processing*, 2015, **54-55**: 41-51
- 6 Gibert X, Patel V M, Chellappa R. Deep multitask learning for railway track inspection. *IEEE Transactions on Intelligent Transportation Systems*, 2017, **18**(1): 153-164
- 7 Giben X, Patel V M, Chellappa R. Material classification and semantic segmentation of railway track images with deep convolutional neural networks. In: Proceedings of the 2015 IEEE International Conference on Image Processing (ICIP), Québec, Canada: IEEE, 2015: 621-625
- 8 Faghieh-Roohi S, Hajizadeh S, Núñez A, Babuska R. Deep convolutional neural networks for detection of rail surface defects. In: Proceedings of the 2016 International Joint Conference on Neural Networks (IJCNN), IEEE, 2016: 2584-2589
- 9 Masci J, Meier U, Ciresan D, et al. Steel defect classification with max-pooling convolutional neural networks. In: Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN), IEEE, 2012: 1-6
- 10 Chen J W, Liu Z Y, Wang H R, Núñez A, Han Z W. Automatic defect detection of fasteners on the catenary support device using deep convolutional neural network. *IEEE Transactions on Instrumentation and Measurement*, 2018, **67**(2): 257-269
- 11 Liu Z G, Wang L Y, Li C J, Yang G J, Han Z W. A high-precision loose strands diagnosis approach for isoelectric line in high-speed railway. *IEEE Transactions on Industrial Informatics*, 2018, **14**(3): 1067-1077
- 12 Zhong J P, Liu Z T, Han Z W, Han Y, Zhang W X. A CNN-based defect inspection method for catenary split pins in high-speed railway. *IEEE Transactions on Instrumentation and Measurement*, 2018
- 13 Yuan Jing, Zhang Yu-Jin. Application of sparse denoising auto encoder network with gradient difference information for abnormal action detection. *Acta Automatica Sinica*, 2017, **43**(4): 604-610 (袁静, 章毓晋. 融合梯度差信息的稀疏去噪自编码网络在异常行为检测中的应用. 自动化学报, 2017, **43**(4): 604-610)
- 14 Tang Xian-Lun, Du Yi-Ming, Liu Yu-Wei, Li Jia-Xin, Ma Yi-Wei. Image recognition with conditional deep convolutional generative adversarial networks. *Acta Automatica Sinica*, 2018,

- 44(5): 855–864
(唐贤伦, 杜一铭, 刘雨微, 李佳歆, 马艺玮. 基于条件深度卷积生成对抗网络的图像识别方法. 自动化学报, 2018, 44(5): 855–864)
- 15 Xin Yu, Yang Jing, Xie Zhi-Qiang. An overlapping semantic community structure detecting algorithm by label propagation. *Acta Automatica Sinica*, 2014, 40(10): 2262–2275
(辛宇, 杨静, 谢志强. 基于标签传播的语义重叠社区发现算法. 自动化学报, 2014, 40(10): 2262–2275)
- 16 Denker J S, Lecun Y. Transforming neural-net output levels to probability distributions. *Advances in Neural Information Processing Systems*, 1991: 853–859
- 17 MacKay D J C. A practical Bayesian framework for back-propagation networks. *Neural Computation*, 1992, 4(3): 448–472
- 18 Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 2014, 15(1): 1929–1958
- 19 Gal Y, Ghahramani Z. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: Proceedings of the 2016 International Conference on Machine Learning, 2016: 1050–1059
- 20 Zheng Wen-Bo, Wang Kun-Feng, Wang Fei-Yue. Background subtraction algorithm with Bayesian generative adversarial networks. *Acta Automatica Sinica*, 2018, 44(5): 878–890
(郑文博, 王坤峰, 王飞跃. 基于贝叶斯生成对抗网络的背景消减算法. 自动化学报, 2018, 44(5): 878–890)
- 21 Fu J, Zheng H, Mei T. Look closer to see better: recurrent attention convolutional neural network for fine-grained image recognition. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4438–4446
- 22 Wang F, Jiang M Q, Qian C, Yang S, Li C, Zhang H G, et al. Residual attention network for image classification. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 3156–3164
- 23 He K M, Gkioxari G, Dollár P, Girshick R. Mask R-CNN. In: Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017: 2961–2969
- 24 Lin H, Shi Z, Zou Z. Fully convolutional network with task partitioning for inshore ship detection in optical remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 2017, 14(10): 1665–1669
- 25 Chen L C, Zhu Y K, Papandreou G, Schroff F, Adam H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the 2018 European Conference on Computer Vision (ECCV), 2018: 801–818
- 26 Han Jiang-Hong, Qiao Xiao-Min, Wei Xing, Lu Yang. Down-hole track detection method based on spatial convolutional neural network. *Journal of Electronic Measurement and Instrumentation*, 2018, 32(12): 34–43
(韩江洪, 乔晓敏, 卫星, 陆阳. 基于空间卷积神经网络的井下轨道检测方法. 电子测量与仪器学报, 2018, 32(12): 34–43)
- 27 Shi Zeng-Lin, Ye Yang-Dong, Wu Yun-Peng, Lou Zheng-Zheng. Crowd counting using rank-based spatial pyramid pooling network. *Acta Automatica Sinica*, 2016, 42(6): 866–874
(时增林, 叶阳东, 吴云鹏, 娄铮铮. 基于序的空间金字塔池化网络的人群计数方法. 自动化学报, 2016, 42(6): 866–874)
- 28 Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. Semantic image segmentation with deep convolutional nets and fully connected CRFs, arXiv preprint arXiv: 1412. 7062, 2014
- 29 Zhang Hui, Jin Xia-Ting, Wu Q. M. Jonathan, He Zhen-Dong, Wang Yao-Nan. Automatic visual detection method of railway surface defects based on curvature filtering and Improved GMM. *Chinese Journal of Scientific Instrument*, 2018, 39(4): 181–194
(张辉, 金侠挺, Wu Q M Jonathan, 贺振东, 王耀南. 基于曲率滤波和改进 GMM 的钢轨缺陷自动视觉检测方法. 仪器仪表学报, 2018, 39(4): 181–194)
- 30 Luo Xiao-Fei, Xu Jun, Chen Jia-Mei. A deep convolutional network for pixel-wise segmentation on epithelial and stromal tissues in histologic images. *Acta Automatica Sinica*, 2017, 43(11): 2003–2013
(骆小飞, 徐军, 陈佳梅. 基于逐像素点深度卷积网络分割模型的上皮和间质组织分割. 自动化学报, 2017, 43(11): 2003–2013)
- 31 Chollet F. Xception: deep learning with depthwise separable convolutions. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1251–1258
- 32 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. In: Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431–3440
- 33 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, 2016: 779–788
- 34 Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Proceedings of the 2015 International Conference on Medical image Computing and Computer-assisted Intervention, Springer, Cham, 2015: 234–241
- 35 Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481–2495
- 36 Zhao H S, Shi J P, Qi X J, Wang X G, Jia J Y. Pyramid scene parsing network. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2881–2890



金侠挺 湖南大学电气与信息工程学院硕士研究生. 2017 年获得长沙理工大学学士学位. 主要研究方向为机器学习, 深度学习, 视觉检测.

E-mail: xtchin@hnu.edu.cn

(**JIN Xia-Ting** Master student at the College of Electrical and Information Engineering, Hunan University. He received his bachelor degree from Changsha University of Science and Technology in 2017. His research interest covers machine learning and deep learning.)



王耀南 中国工程院院士, 湖南大学电气与信息工程学院教授. 1995 年获得湖南大学博士学位. 主要研究方向为机器人学, 智能控制和图像处理. 本文通信作者.

E-mail: yaonan@hnu.edu.cn

(**WANG Yao-Nan** Academician at Chinese Academy of Engineering, professor at the College of Electrical and Information Engineering, Hunan University. He received his Ph.D. degree from Hunan University in 1995. His research interest covers robotics, intelligent control, and image processing. Corresponding author of this paper.)



张辉 长沙理工大学副教授. 2012 年获得湖南大学博士学位. 主要研究方向为工业机器视觉, 数字图像处理.

E-mail: zhanghuihy@126.com

(**ZHANG Hui** Associate professor at the College of Electrical and Information Engineering, Changsha University of Science and Technology. He received his Ph.D. degree from Hunan University in 2012. His research interest covers industrial machine vision and digital image processing.)



刘理 湖南大学博士研究生. 2006 年获得东南大学硕士学位. 主要研究方向为机器人视觉测量, 路径规划及智能控制.

E-mail: liuli@hnu.edu.cn

(**LIU Li** Ph.D. candidate at the College of Electrical and Information Engineering, Hunan University. He received his bachelor degree in measurement and control technology and instrument from Southeast University in 2006. His current research interest covers robot vision measurement, robot path planning and intelligent control.)



钟杭 湖南大学博士研究生. 2013 年和 2016 年分别获得湖南大学学士学位和硕士学位. 主要研究方向为机器人控制, 视觉伺服和路径规划.

E-mail: zhonghang@hnu.edu.cn

(**ZHONG Hang** Ph.D. candidate at the College of Electrical and Information Engineering, Hunan University. He received his master and bachelor degrees in automation from Hunan University in 2013 and 2016, respectively. His current research interest covers robotics control, visual servo and path planning.)



贺振东 郑州轻工业大学副教授. 2016 年获得湖南大学博士学位. 主要研究方向为机器视觉, 机器学习.

E-mail: hezhendong_itl@163.com

(**HE Zhen-Dong** Associate professor at the College of Electrical and Information Engineering, Zhengzhou University of Light Industry. He received his Ph.D. degree from Hunan University in 2016. His current research interest covers machine vision, robotics, and machine learning.)