计算机科学与探索
Journal of Frontiers of Computer Science and Technology
ISSN 1673-9418,CN 11-5602/TP

《计算机科学与探索》网络首发论文

题目：结构化区域全卷积神经网络的钢轨扣件检测方法
作者：蒋欣兰
网络首发日期：2020-10-19
引用格式：蒋欣兰．结构化区域全卷积神经网络的钢轨扣件检测方法．计算机科学与探索. https://kns.cnki.net/kcms/detail/11.5602.TP.20201019.1151.002.html

计算机科学与探索
Journal of Frontiers of Computer Science and Technology doi: 10.3778/j.issn.1673-9418.2009065

结构化区域全卷积神经网络的钢轨扣件检测方法*

蒋欣兰 [1,2+]

1. 中国社会科学院大学 计算机教研部，北京 102488
2. 中国社会科学院大学 计算社会科学研究中心，北京 102488
+ 通信作者 E-mail: jxlzgl@sina.com

摘 要：现有的深度学习模型很难满足高速检测的实时性，针对性地提出了一种结构化区域全卷积神经网络（SR-FCN）。为了满足高速综合巡检车实时检测的任务，考虑到轨道图像中钢轨、扣件、轨道板等设施位置相对固定，其位置分布可以构成轨道场景特有的结构化特征，因此设定了结构化检测区域，将一幅图像中扣件小目标的检测转化为一整块具有固定结构的大目标区域检测，将扣件小目标的检测问题转化为结构化区域的定位问题，可加快网络的训练收敛速度，减少候选区域的生成个数，从而大幅提高检测速度。将铁路轨道的结构化先验信息融合到深度学习网络的各个过程中，有效提高了定位精度，保证了检测的鲁棒性。实验室离线分析和现场在线检测的结果表明，所提出的 SR-FCN 网络分别获得了 99.99%和 99.84%的检测精度，同时还保持了较快的检测速度，可以满足 350km/h 的实时检测要求。

关键词：目标检测；深度学习；扣件；结构化场景；高速巡检
文献标志码：A **中图分类号：**TP391

蒋欣兰. 结构化区域全卷积神经网络的钢轨扣件检测方法[J]. 计算机科学与探索
JIANG X L. Rail Fastener Detection Method Based on Structured Region Full Convolution Neural Network[J]. Journal of Frontiers of Computer Science and Technology

Rail Fastener Detection Method Based on Structured Region Full Convolution Neural Network*

JIANG Xinlan[1,2+]

1. Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences, Beijing 102488, China

*The Higher Education Research Program of University of Chinese Academy of Social Sciences under Grant No. GDJY2020016 (2020 年高等教育研究课题); the College Student Research and Career-creation Program of Beijing City under Grant No. 2020bjsdc037 (2020 年北京市大学生科学研究与创业行动计划项目).

2. Research Center for Computational Social Sciences, University of Chinese Academy of Social Sciences, Beijing 102488, China

Abstract: Existing deep learning models are difficult to meet the real-time performance of high-speed detection, and a structured region fully convolutional neural network (SR-FCN) is proposed. In order to meet the task of real-time detection of high-speed comprehensive inspection vehicles, considering that the rails, rail fasteners, track plates and other facilities in the track image are relatively fixed, their position distribution can constitute a unique structural feature of the track scene, structured detection region is set, the detection of small rail fastener targets in an image is converted into a large target region detection with a fixed structure, and the detection of small rail fastener targets is transformed into a positioning problem in a structured region, which can speed up the network training convergence speed, reduce the number of candidate regions generated, thereby greatly improving the detection speed. The structured prior information of the railway track is integrated into the various processes of the deep learning network, which effectively improves the positioning accuracy to ensure the robustness of the detection. The results of laboratory offline analysis and on-site online testing show that the proposed SR-FCN obtains 99.99% and 99.84% detection accuracy, respectively, and maintains a relatively fast detection speed, which can meet the 350km/h real-time detection requirements.
Key words: target detection; deep learning; rail fasteners; structured scenes; high-speed inspection

1 引言

近年来，随着铁路事业的快速发展，我国铁路总里程已达 12.4 万公里。钢轨扣件是轨道上用以连接钢轨和轨枕的轨道基础设施部件，其作用是将钢轨固定在轨枕上，保持轨距并防止钢轨的侧向移动。钢轨扣件发生异常，将使得扣件对钢轨起不到固定作用，对列车的运行安全产生严重的影响。因此，铁路钢轨扣件的服役状态对保障铁路安全运营至关重要，需要对其进行周期性的巡查，及时发现扣件的异常状态。

然而，铁路沿线环境非常复杂，获取到的视频数据的质量参差不齐，关键区域有可能被遮挡或覆盖，扣件的类型多样，这些因素使得现有的基于人工设计特征的视觉检测方法无法满足实际线路的检测要求。近年来，基于深度学习的目标检测技术取得重大突破，极大的提升了目标检测的准确率。但已有的深度学习检测模型大多针对自然场景中的多类物体检测而设计，应用在结构化的特定轨道场景中，可能会出现过拟合的问题。其次，为了满足时速 350km/h 的高速综合检测列车的实时检测需求，对扣件的检测速度提出了极高的要求，而已有的深度学习模型难以满足高速检测的时效性。

在这种背景下，本文提出了一种优化的结构化区域全卷积深度神经网络（structured region fully convolutional neural network, SR-FCN），充分利用轨道的空间结构化信息，将扣件小目标的检测问题转化为结构化区域的定位问题，并通过优化区域提名网络（region proposal network, RPN）的锚点(anchor)遍历个数，极大的提升了扣件的定位速度，并避免了因局部扣件缺失或背景干扰造成的定位错误，提升了检测的鲁棒性。

2 相关工作

图像中包含不同类别的多个目标，目标检测的首要目的是对检测目标进行精确定位，之后在对定位的目标区域进行识别分类。与图像分类相比，对图像中的目标物体进行检测是更为困难，对其建立的深度学习模型也更为复杂。基于深度学习的目标检测总体上分为两派，即基于候选区域生成的 R-CNN（region based convolutional neural networks）[1]系列以及基于回

归方法的（无需区域提名）YOLO（you only look once）[2]、SSD（single shot detector）[3]系列。

对于基于候选区域生成的检测算法，目标检测的第一步是生成候选区域（region proposal），也就是找出可能的感兴趣区域（region of interest, ROI）。常见的区域生成方法有：1）滑动窗口。滑动窗口本质上就是穷举法，利用不同的尺度和长宽比把所有可能的大大小小的块都穷举出来。2）规则块。在穷举法的基础上进行了一些剪枝，只选用固定的大小和长宽比。3）选择性搜索。从机器学习的角度来说，前面的方法可以取得较好的召回率，但是精度差强人意，所以问题的核心在于如何有效地去除冗余候选区域。其实冗余候选区域大多是发生了重叠，选择性搜索利用这一点，自底向上合并相邻的重叠区域，从而减少冗余。最经典的基于候选区域的深度学习目标检测模型 R-CNN 由 Ross Girshick 等人提出，该模型首先使用选择性搜索（selective search）这一非深度学习算法来定位待分类的候选区域，然后将每个候选区域输入到卷积神经网络中提取特征，接着将这些特征输入到线性支持向量机中进行分类，并在 PASCAL VOC 数据集上取得了比传统算法高约 0.220 的平均正确率，为之后的基于卷积神经网络的目标检测模型构建奠定了基础。

R-CNN 模型需要对图像中所有的候选区域窗口都进行特征提取，这必然导致特征提取的时间耗费巨大。微软亚洲研究院（MSRA）的 K. He[4]提出的空间金字塔池化网络层（spatial pyramid pooling networks，SPP-Net ）针对 R-CNN 时间消耗较大的缺陷进行了改进，仅对输入图像进行一次卷积计算，大大提高了算法的执行效率。

Ross Girshick 也意识到了 R-CNN 速度慢的问题，提出了一种改进的方法 Fast R-CNN[5]。与 R-CNN 中的卷积神经网络层相比，Fast R-CNN 提出了针对性的兴趣域池化层（ROI Pooling）技术对最后一个池化层进行了改进。该层的作用与 SPP-Net 中的空间金字塔池化层相似，对任意大小的输入都输出固定维数的特征向量，也是仅在最后一层卷积层对候选区域的卷积特征进行 ROI Pooling。此外，Fast R-CNN 同时对用于目标定位和分类的两个全连接层进行训练，实现了目标定位与检测分类的一体化。

Ren Shaoqing 等提出了一种多阶段的 R-CNN 网络训练训练算法，称作 Faster R-CNN[6]。Faster R-CNN 基于 R-CNN 网络前几层卷积层提取的特征对检测目标进行定位，且网络的构建利用 GPGPU（general-purpose computing on graphics processing units）来实现，从而大幅降低了整个网络构建的时间消耗，检测所需时间约为原来的 1/10。

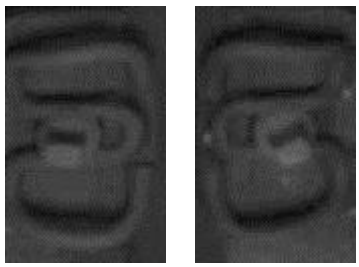
R-CNN 系列方法是目前主流的目标检测方法，但是速度上并不能满足实时的要求。YOLO[2][7][8][9] 等基于回归的一类方法慢慢显现出其重要性，这类方法使用了回归的思想，对于给定的输入图像，直接在图像的多个位置上回归出检测目标的边框以及类别。YOLO 整个过程非常简单，将目标检测任务转换成一个回归问题，不需要执行耗时的区域生成来定位目标，直接回归便完成了位置和类别的判定，大大提高了检测的速度，每秒钟可以处理 45 张图像。但同时由于取消了候选区域生成机制，也导致 YOLO 的检测精度并不是很理想。SSD 将 YOLO 的回归思想以及 Faster R-CNN 的锚点机制有机结合，使用全图各个位置的多尺度区域特征进行回归，既保持了 YOLO 快速检测的特性，也保证了 Faster R-CNN 窗口提取的精准性。

R-FCN（region-based fully convolutional networks）[10]是基于 Faster-RCNN 的改进，将耗时的全连接层转化为卷积操作构成全卷积网络，并引入区域敏感度的概念，显著提高了目标检测的精度和速度。

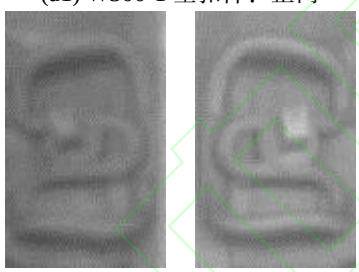
3 存在的问题

轨道巡检车采集的图像中轨道场景具有固定的空间结构，扣件类型多样且形状差异性较小。此外，高速巡检对扣件的检测速度也提出了极高的要求。下面从扣件定位的类型、速度和精度三个方面阐述现有深度学习检测模型在扣件检测中面临的问题和挑战。

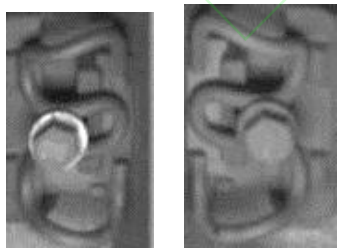
（1） 扣件定位的类型



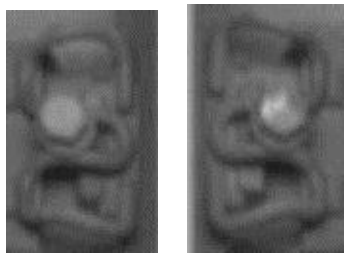
(a1) W300-1 type rail fastener: forward driving
(a1) W300-1 型扣件：正向



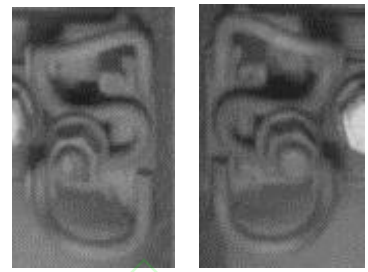
(a2) W300-1 type rail fastener: reverse driving
(a2) W300-1 型扣件：反向



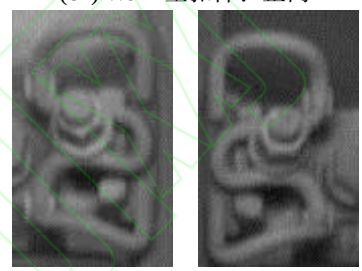
(b1) V type rail fastener: forward driving
(b1) V 型扣件：正向



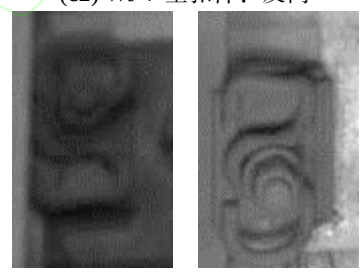
(b2) V type rail fastener:reverse driving
(b2) V 型扣件：反向



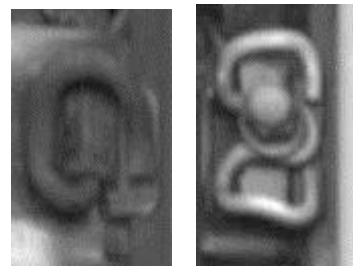
(c1) WJ-7 type rail fastener: forward driving
(c1) WJ-7 型扣件：正向



(c2) WJ-7 type rail fastener: reverse driving
(c2) WJ-7 型扣件：反向



(d1) Ballast line rail fastener: forward driving
(d1) 有砟线路扣件：正向



(d2) Ballast line rail fastener：reverse driving
(d2) 有砟线路扣件：反向

Fig.1 Rail fastener type

图 1 扣件类型

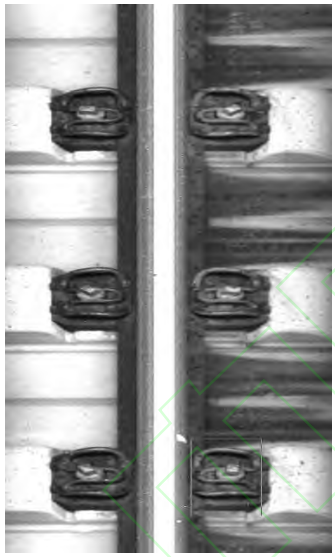
如图 1 所示为线路中几种典型的扣件类型，其中无砟线路包括 W300–1 型、Ⅴ型、WJ–7 型扣件，有砟线路也有多种扣件类型（与无砟扣件相比较小）。此外，由于成像设备与钢轨平面存在一个拍摄角度，从而造成正反向行车时扣件成像不同。如

图所示，正向行车时，扣件弹条的上部为曲折状，下部为圆弧状；而反向行车时，则正好相反。

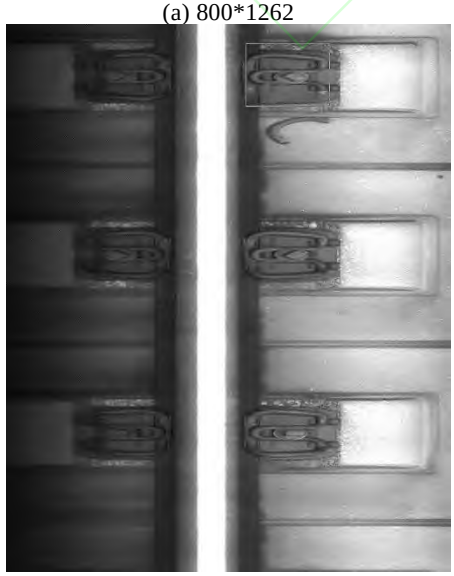
由于线路中扣件类型多样，且区分正反向，因此基于深度学习的方法定位扣件时需综合考虑扣件类型的多样性。

（2） 扣件定位的速度

如图 2 所示，不同检测车采集的轨道图像大小不同。检测算法应在不同分辨率的图像上取得近似的时间耗费。若一帧图像空间采样为 2 米，要满足 350km/h 速度下的扣件实时定位。则检测速度要求能够达到 49 帧/秒，即 20ms/帧的检测耗时。



(a) 800*1262
(a) 800*1262



(b) 1024*1262
(b) 1024*1262

(c) 1024*1011
(c) 1024*1011

Fig.2 Track image size

图 2 轨道图像大小

（3） 扣件定位的精度

以 100 公里的轨道巡检数据为例，总共存在约 600000（六十万）个扣件。综合当前巡检系统扣件定位的精度以及用户人工复核的主观要求，除去道岔、区段、联络线区段，无砟正线错误定位的扣件不应超过 100 个/100 公里，即定位的准确率要求>99.98%。

现有的深度学习检测模型都是为了解决自然场景中的多类目标定位问题提出的，轨道场景结构相对比较固定，局部变化较小，针对复杂自然场景的深度网络模型在训练过程中容易过拟合；检测过程中没有充分利用场景的固定结构化空间信息，抗干扰能力差。

针对以上存在的问题，本文结合轨道扣件的结构化分布特征，提出了一种改进的区域全卷积深度神经网络，并根据检测任务的特性对卷积网络、候选区域生成网络、区域敏感得分图等进行了相应地优化调整，在保证检测精度的同时，极大地提升了目标检测的速度，可满足高速动态检测的需求。

4 方法与论证

如图 3 所示，基于深度学习的检测过程分为两个阶段，即“离线训练”和“在线检测”。首先从大量的轨道图像中利用模板匹配的方法进行样本自动标注，构建用于学习的大数据样本集，输入深度网络中进行离线训练和调试得到网络模型参数；利用训练得到的模型参数初始化深度网络的模型参数，赋予网络目标检测的能力，将单幅待检测的轨道图像输入参数初始化的深度网络中实现目标的在线实时检测。

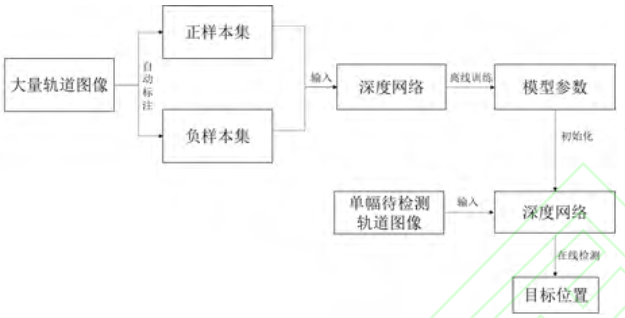


Fig.3 The basic process of target detection based on deep learning

图 3 基于深度学习的目标检测基本流程

本文旨在解决高速行车下的钢轨扣件实时动态检测问题，结合检测目标的空间分布特征，对深度学习检测模型中的各个环节进行了改进和优化，主要包括以下几个方面：

（1）轨道图像中钢轨、扣件、轨道板等设施相对固定的位置分布构成了轨道场景特有的结构化特征，本文将一幅图像中多个扣件小目标的检测转化为一整块具有固定结构的大目标区域检测，可加快网络的训练收敛速度，减少候选区域的生成个数，从而提高检测的速度；

(2)针对具有相似结构的大目标检测区域，检测目标形状相似，仅在空间位置上存在一定的变化，为了防止训练过拟合，本文使用 ResNet-18 作为卷积层的网络结构；

(3)本文构造了轨道场景各设施部件对应的位置敏感得分图，将目前最快的 R-FCN 深度网络与轨道场景的结构化信息有机结合，提出结构化区域深度全卷积网络（structured region fully convolutional neural network，SR-FCN），可有效解决高速实时检测任务中的速度瓶颈，并提高目标检测的精度以及抗干扰能力；

(4)本文针对轨道场景的固定结构对候选区域生成网络（RPN）进行了改进，通过约束锚点的遍历范围，并限定候选窗口尺度变化比例，减少生成候选区域的个数，进一步加快目标检测的速度；

(5)本文根据检测目标的位置分布特征(如轨道图像中扣件呈”田”字形分布), 对深度网络的损失函数进行空间分布正则化，进一步保证了扣件目标检测的精度和容错能力。

4.1 轨道场景结构解析

无论是 Faster-RCNN 还是 R-FCN 深度模型，在图像中进行目标定位都是基于窗口滑动搜索的方法生成检测候选区域，而目前轨道巡检系统采集的轨道图像最小尺寸为 800×1230，扣件区域尺寸为 80×128，直接使用滑动窗口法进行穷举搜索将会严重影响系统效率。事实上，铁路轨道图像中至少包含了 7 个固定结构的先验知识：(1)每帧轨道图像中只包含一条钢轨；(2)钢轨总是与图像的 x 轴垂直，并且钢轨的两条边界是平行的；（3）每幅图像高度方向的空间采样距离为 2m 且误差小于 2mm； (4)钢轨的宽度是固定的像素值；(5)扣件区域总是在钢轨边界的两侧对称分布，并且扣件区域的尺寸是固定的； (6) 相邻扣件承轨台或轨枕沿竖直方向的间距相对固定；（7）每幅图中包含 6 个完整的扣件，且呈“田”字型分布。

图 4 展示了轨道图像中的位置先验信息，钢轨的宽度为 60 像素，扣件区域的宽度为 80 像素，扣件横向间隔约为 55-65 像素，纵向间隔约为 275-315 像素。

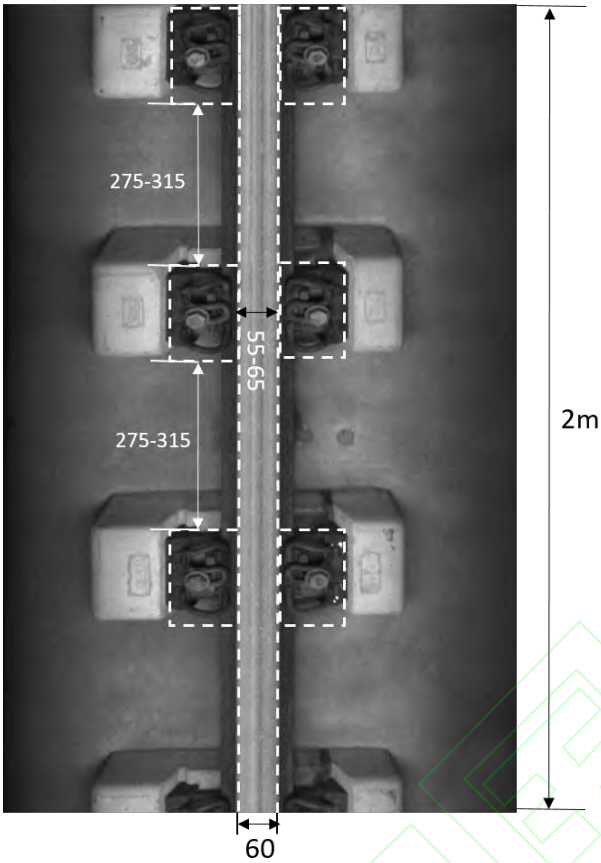


Fig.4 Structured prior information of track scene

图 4 轨道场景结构化先验信息

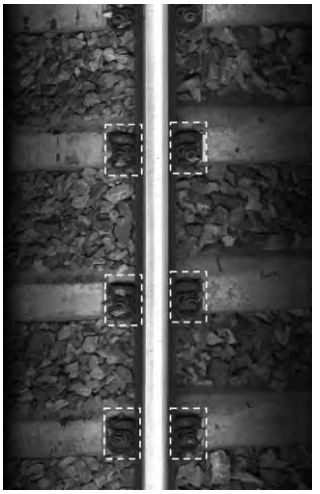
本文充分利用场景中已知的先验信息，将其融合到深度网络中的样本构造，候选区域生成，网络构造以及损失函数约束等各个过程，大幅减小了扣件候选区域的范围，提高了检测效率并保证了检测精度。

4.2 训练样本构造

如图 5 所示，为了使用深度学习网络检测钢轨的扣件位置，首先需要在轨道图像上对检测的目标进行大量的人工标注工作。以往的标注方法如图 5(a)所示，直接对轨道图像中的 6 个扣件检测目标进行标注作为训练样本，利用深度学习网络进行模型训练。实际检测环境中，扣件的形态易受道砟覆盖、光照环境变化、扣件状态异常等多种不确定因素的影响，对扣件样本的多样性有很高的要求。此外，标注的扣件样本作为典型的小目标样本，候选区域生成的数量更多，检测较为耗时，且由于卷积层的多次池化操作使得网络对小目标的检测不敏感。

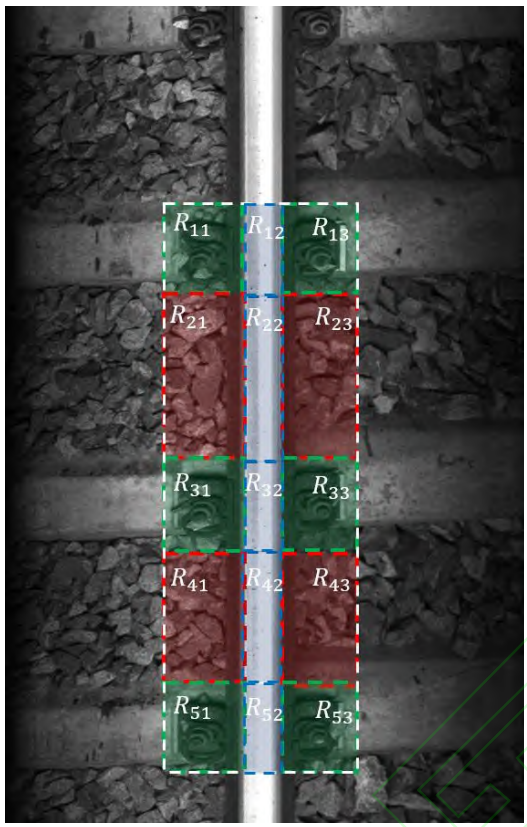
本文提出了一种大目标结构化区域自动标注方法，根据上一节中场景结构解析的结果，将扣件、道砟、钢轨等多个子区域构成整体的大目标结构化检测区域。这一改进首先将多个小目标检测任务转化为单一的大目标检测，提高了检测速度；其次，对该大目标的检测可充分利用各个子区域之间相对固定的位置和形状约束关系，可有效提高检测的抗干扰能力。

如图 5(b)所示为标注的大目标结构化区域，可划分为 15 个子区域。其中，R_{11}、R_{13}、R_{31}、R_{33}、R_{51}、R_{53}为扣件子区域；R_{21}、R_{23}、R_{41}、R_{43}为道砟子区域；R_{12}、R_{22}、R_{32}、R_{42}、R_{52}为钢轨子区域。这里需要注意的是，R–FCN 网络将检测目标等间隔的划分为$K\times K$（默认为3×3）个子区域，不同与 R–FCN 的划分方式，本文提出的 SR–FCN 网络将标记的大目标检测区域划分为5×3个子区域，且各子区域的大小和间距按照轨道场景的结构化先验来进行初始化。



(a) Small target rail fastener labeling

(a) 小目标扣件标注



(b) Large target structured area labeling

(b) 大目标结构化区域标注

Fig.5 Deep learning sample labeling method

图 5 深度学习样本标注方法

(a) Positive sample of ballasted track

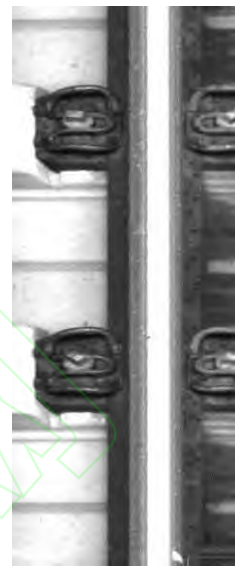
(a) 有砟轨道正样本

(b) Negative sample of ballasted track

(b) 有砟轨道负样本

(c) Positive sample of ballastless track

(c) 有砟轨道正样本



(d) Negative sample of ballastless track

(d) 有砟轨道负样本

Fig.6 Sample automatic label template

图 6 样本自动标注模板

为了提高样本标注的效率，本文提出了基于模板匹配的样本自动标注方法，具体流程如下：（1）首先手工标记目标检测区域以及非目标检测区域（如图 6 所示）分别作为正、负样本模板添加到对应的正、负模板库中。为了增加样本的多样性，定义与检测目标区域重合度（即像素交并比 intersection over union，IOU）大于 80% 的区域可作为正样本，而重合度低于 50%的作为负样本；

（2）然后，对每一帧轨道图像，利用滑动窗口法从轨道图像中提取子窗口，提取子窗口的 HoG（histogram of oriented gradient）特征与正、负数模板库中每个模板计算二者之间的相似度，按照相似度从高向低选择 K 个模板，利用 K–NN(K– Nearest Neighbor)分类器对子窗口所属类别进行投票，在正模版库中得分较高的子窗口被自动标注为正样本；而负模板库中得分较高的子窗口被自动标注为负样本；

（3）对自动标注的结果进行人工复核，去除错误的生成样本，完成样本的清洗。

4.3 网络模型构建

深度学习的基本工作原理可概括如下：

$$g = \mathrm{A}f$$

这里，f 表示输入图像；g 表示检测结果；A 为一变换矩阵，表征输入与输出之间的对应关系。

则对于网络训练而言，其本质是基于大量的输入数据 f 以及事先标记的输出结果即样本 g，通过迭代逼近的方法估算二者之间的变换矩阵 A，当满足迭代次数条件或者网咯误差达到预测的阈值以下，则认为求得了近似于 A 的变换矩阵 $\hat{\mathrm{A}}$，$\hat{\mathrm{A}}$ 也称作通过训练得到的网络模型参数。

在得到网络模型参数 $\hat{\mathrm{A}}$ 后，目标检测的过程可记作

$$g_0 = \hat{\mathrm{A}} f_0$$

即对于一幅待检测的图像 f_0，利用训练得到的模型参数与其变换运算，即可得到目标检测的结果 g_0。

如图 7 所示为 SR–FCN 网络的组织结构，网络的训练和检测过程可以表述如下：

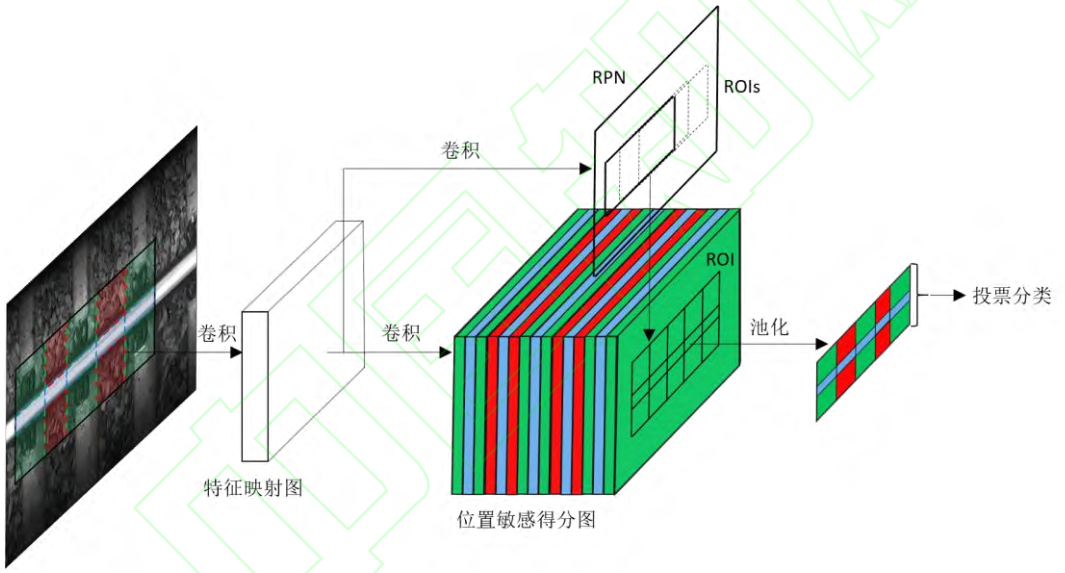


Fig.7　SR-FCN network structure

图 7　SR-FCN 网络结构

网络的训练过程如下：

（1）按照 4.2 中的方法构造训练样本，即构造用于训练的训练图像 f 并预标记对应的检测结果 g，用于学习模型参数 $\hat{\mathrm{A}}$。为了保证网络的训练效果，需要对不同分辨率的轨道图像进行相应的尺度调整，统一成像的空间分辨率大小，即不同图像中的单个像素应对应相同的空间采样间隔；

（2）网络的训练过程和参数调节过程和 R–FCN 类似，需要训练的模型参数由三部分组成，即用于生成特征映射图的多组卷积核 C，RPN 网络总用于生成候选区域的多组卷积核 R，以及用于生成位置敏感图得分的多组卷积核 P，这三部分参数共同作用构成了训练模型参数 $\hat{\mathrm{A}}$，训练的本质就是根据标记的样本数据来求得网络中各部分的模型参数；

（3）选择模型调整优化策略如随机梯度下降法（stochastic gradient descent，SGD）对模型参数进行不断的迭代逼近，直到迭代次数达到预设的次数或网络的训练误差小于预设阈值。

在学习得到网络模型参数后，各部分的卷积核参数 C、R、P 已知，则网络的检测过程可描述如下：

（1）选择一张待检测的轨道图像，并对该图像进行相应的尺度调整、灰度归一化等预处理操作。

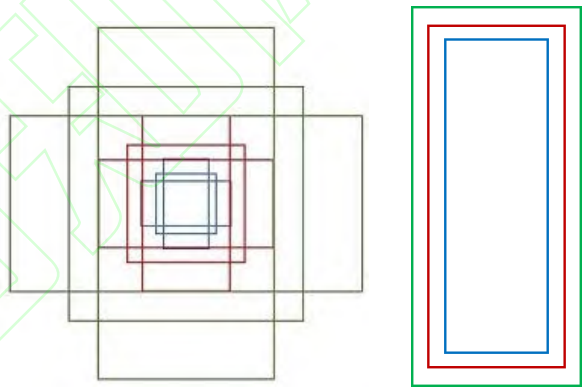
（2）将预处理后的图片送入一个预训练好的分类网络中，经过特征卷积层 C 生成特征映射图。由于检测目标具有固定的结构化信息，且检测类型单一，为了防止网络过拟合并提高检测速度，网络层数不宜过深，这里使用 VGG16、ResNet-18 [11,12] 等网络结构作为特征学习的卷积模型。

（3）利用训练好的 RPN 网络模型参数 P 生成目标检测的候选区域，并在候选区域上利用训练好的网络模型参数 P 生成目标检测的位置敏感得分图。需要说明的是，与 R-FCN 类似，改进的 SR-FCN 网络在预训练网络的最后一个卷积层获得的特征图上也存在 3 个分支，第 1 个分支就是在该特征图上面利用区域候选网络（RPN）生成候选区域，获得相应的感兴趣检测区域（ROI）；第 2 个分支就是在该特征图上获得一个 5×3×(C+1) 维的位置敏感得分映射（position-sensitive score map）用来进行分类，这里 C 表示检测目标的类别数目，即训练过程中将大目标结构化区域标记为几类；第 3 个分支就是在该特征图上获得一个 5×3×4 维的位置敏感得分映射（每个子区域的位置可记为一个四元组 (x,y,w,h)，15 个子区域一共 5×3×4 个位置得分映射），用来回归调整每个子区域的位置。

（4）基于 R-FCN 网络的 ROI 池化方法以及投票分类规则，在 5×3×(C+1) 维的位置敏感得分映射和 5×3×4 维的位置敏感得分映射上面分别执行位置敏感的 ROI 池化操作（Position-Sensitive ROI Pooling，这里使用的是平均池化操作），并经区域投票和局部回归获得对应的类别和位置信息。

4.4 候选区域生成网络改进

原始的锚点生成规则在特征图的每个像素点上都生成 9 个不同尺度和不同长宽比的候选区域，如图 8(a)所示。本文针对结构化检测区域尺度和高宽比相对比较固定的特性，对锚点生成规则进行了改进如图 8(b)所示，每个像素点上仅生成三个高宽比固定且尺度轻微缩放的候选区域，大幅较少了生成候选区域的个数，进一步提高了目标检测的速度。



(a) Original anchor generation rules　(b) Improved anchor generation rules

(a) 原始锚点生成规则　(b) 改进的锚点生成规则

Fig.8 Anchor generation rules

图 8 锚点生成规则

4.5 结构正则化损失函数定义

基于轨道场景中目标分布的结构化信息，可以引入结构保持正则化项 $\lambda_2\left[c^*>0\right]L_{sr}(h,h^*)$ 来加快网络收敛，避免训练迭代陷入局部最小解。提出的损失函数定义如下：

$$L\left(s,t_{x,y,w,h},h_{l_t,r_t,l_m,r_m,l_b,r_b}\right)=L_{cls}(s_{c^*})+\lambda_1\left[c^*>0\right]L_{reg}(t,t^*)+\lambda_2\left[c^*>0\right]L_{sr}(h,h^*)$$

上式为本文定义的结构正则化损失函数，包括一个目标分类损失计算 L_{cls}，一个位置回归损失计算 L_{reg}，以及一个结构保持损失计算 L_{sr}。λ_1 和 λ_2 用来平衡三者的重要度；s_{c^*}、t^*、h^* 分别为用于训练的样本标签；$c^*>0$ 表示检测的目标非背景；$s,t_{x,y,w,h}$，$h_{l_t,r_t,l_m,r_m,l_b,r_b}$ 表示训练的输入数据。

5 实验与分析

5.1 实验环境与数据

本文的实验环境和系统配置如下：

（1）硬件配置: Intel Xeon@ 2.40 GHz×28 + NVIDIA Geforce Titan X×4 + 256GB 内存

（2）操作系统：Ubuntu 16.04 LTS

（3）深度学习框架: CUDA 8.0 + Anaconda Python 2.7 + Caffe

实验数据来自于检测车在全国各有砟、高铁线路采集的轨道巡检图像，用于深度学习网络训练、测试以及 SR-FCN 检测效果的验证，样本图像组成结构如表 1 所示。

Table 1 Inspection sample image

表 1 巡检样本图像

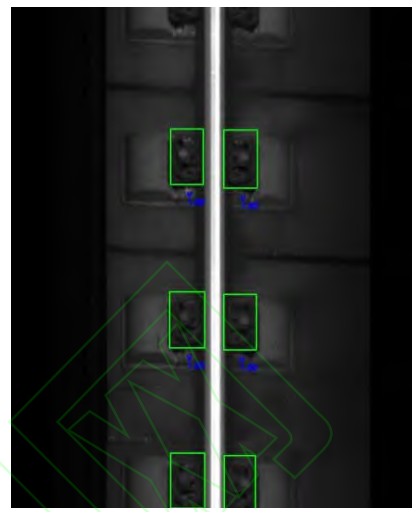
线路类型	训练样本集个数（含验证）	测试样本集个数
有砟线路	20000	2000
高铁线路	40000	4000
总数	60000	6000

5.2 网络训练参数

考虑到网络中主要采用了 ReLU 作为激活函数，因此采用 Kaiming 初始化方法对网络进行初始化。使用随机梯度下降(SGD)模型优化方法，初始学习率(learning rate)设为 0.01，动量参数(momentum)设为 0.9，权值衰减(weight decay)设为 0.0005。每当图片在 Validation 数据集的目标函数值相比前一次迭代没有下降的时候，将学习率减小为原来的 10%，即学习率设为 0.001； 批处理大小(batch size)设为 256, 训练集迭代次数(epochs)设为 100。

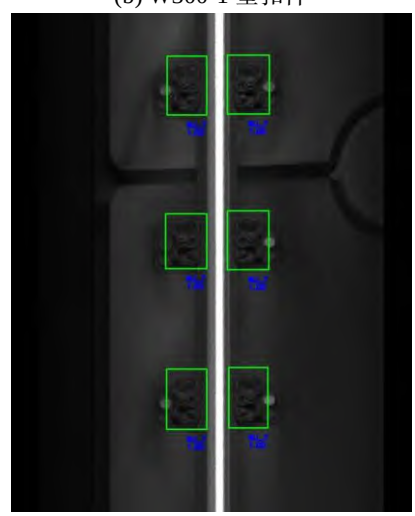
5.3 实验室静态试验结果

基于 SR-FCN 的扣件目标检测的部分结果如图 9 所示，针对不同轨道场景的不同扣件类型，扣件区域的置信度得分都接近 1，检测精度高，场景适应能力强。

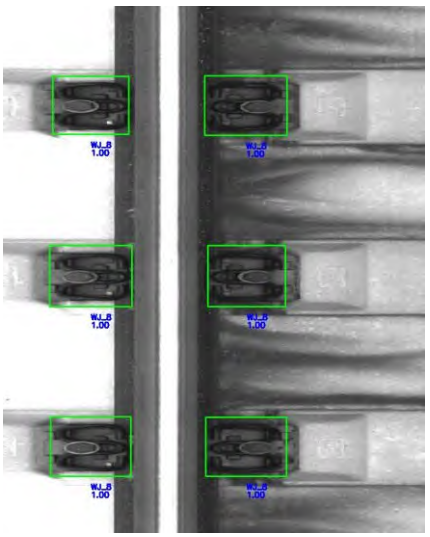


(a) V type rail fastener
(a) V 型扣件

(b) W300-1 type rail fastener
(b) W300-1 型扣件



(c) WJ-7 type rail fastener
(c) WJ-7 型扣件



(d) WJ-8 type rail fastener
(d) WJ-8 型扣件

Fig.9 Test results of different types of rail fasteners

图 9 不同类型扣件的检测结果

利用不同的深度学习网络对表 1 中的 6000 幅图像进行测试，对结果进行统计对比各方法的检测的准确率和检测速度，如表 2 所示。定义检测成功率($P_{d)}$来评价方法的准确率，成功率定义为

$$P_{\mathrm{d}}=\frac{area(R_t \cap R_g)}{area(R_t \cup R_g}$$

这里，R_t 表示检测结果的边缘矩形框，R_g 表示手工标注的目标真实位置矩形框，如果 SR>0.8，则认为该图像的检测结果是正确的。最后统计所有的 N 帧测试图像中追踪成功的图像总数，记为 T，则追踪准确率记为 T/N。

Table 2 Comparison of rail fastener detection results of different deep learning models

表 2 不同深度学习模型的扣件检测结果对比

网络模型	检测耗时(ms)	检测速度（km/h）	检测成功率
Faster-RCNN ResNet50	170	42	99.64%
R-FCN ResNet101	150	48	99.89%
SSD Inception v2	61	118	98.65%
YOLO v3	55	131	99.98%
SR-FCN ResNet18	19	379	99.99%

本文提出的 SR–FCN 网络不仅可以适应多种场景下的目标定位检测，而且仍然能够保持很高的检测速度。如表 3 所示，提出的网络模型在单幅轨道图像上执行目标定位耗时为 38ms，对应于可以满足 360km/h 速度下的实时检测。这里需要说明的是，对于钢轨扣件而言，100 公里范围内大约分布有 60 万个扣件，扣件定位算法对检测成功率的要求很高，如 100 公里线路仅 1%的误报将产生 6000 个错误定位的结果，将严重影响后续的识别分析。因此，对于扣件定位算法而言，0.1%的检测精度提升也具有很强的实际意义。

5.4 线上动态试验结果

我们将本文提出的模型部署在综合巡检车上，并在 4 条实际运营线路进行动态实验论证。

本试验采用检出率（detection rate，DR）、误检率（detection error rate，DER）两个评价指标来评价轨道扣件区域定位方法的有效性和可靠性。检出率用于评价轨道扣件区域定位方法的有效性，即定位出的有效扣件区域的数量与真实扣件区域的数量的比例，检出率越高，表示方法的有效性越强。误检率用于评价轨道扣件区域定位方法的可靠性，即定位出的所有扣件区域中无效的比例，误检率越低，表示方法的可靠性越强。检出率和误检率的计算公式如下：

$$DR=\frac{N_{Valid}}{N_{GT}}\times 100\%$$

$$DER = \frac{N_{Invalid}}{N_{Invalid} + N_{Valid}} \times 100\%$$

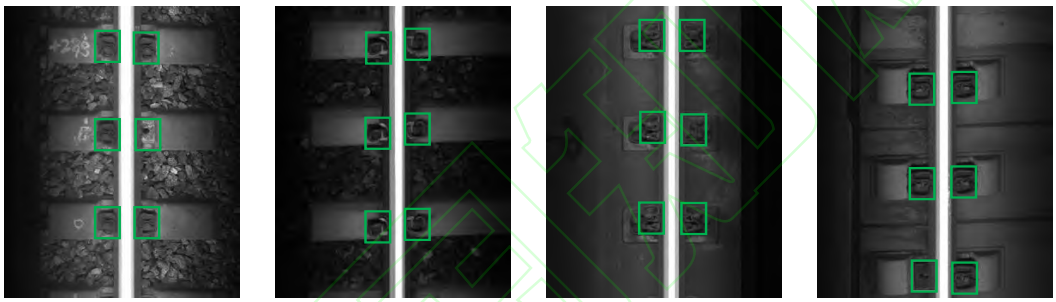
式中，N_{Valid} 表示定位出的扣件区域中有效的数量，$N_{Invalid}$ 表示定位出的扣件区域中无效的数量，N_{GT} 表示真实扣件区域的数量。

如表 3 所示，本试验中，在 4 条铁路线路的平均检出率达到 99.84%，平均误检率低至 0.62%，试验结果表明，虽然相对于静态试验结果，检出率有所下降，但对于训练样本集之外的新线路轨道图像，本文提出的方法具有令人满意的检出率和泛化能力，可以满足现场应用需求，各线路轨道扣件区域定位结果如图 10 所示。

Table 3　Online dynamic test comparison results

表 3　线上动态试验对比结果

线路名	检出率(%)	误检率(%)
1#	99.53	0.82
2#	99.95	0.76
3#	99.87	0.52
4#	99.99	0.37
平均	99.84	0.62



(a) 1#railway　(b) 2#railway　(c) 3#railway　(d) 4#railway

(a) 1#铁路　(b) 2#铁路　(c) 3#铁路　(d) 4#铁路

Fig.10　Example map of the location results of the track rail fastener area

图 10　各线路轨道扣件区域定位结果示例图

6　结论

本文提出了一种结构化区域全卷积神经网络结构（SR–FCN），实现了结构化轨道场景中多模态扣件目标的快速、准确、鲁棒的检测。首先，构建了一种结构化场景下基于区域推理的学习训练样本构造和标记方法，将场景中的多个离散小目标转化为固定几何结构约束的大目标，大幅提高网络训练的效率和目标检测的速度。在此基础上，提出了一种基于几何位置先验的候选区域生成网络，进一步减少了目标定位候选区域的数量，有利于提高目标检测的速度。最后，定义了一种结构正则化损失函数，根据检测目标的位置分布特征，对深度网络的损失函数进行空间分布正则化，进一步保证了扣件目标检测的精度和容错能力。

References:

[1]　Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Jun 23-28, 2014.Washington:IEEE Computer Society, 2014: 580-587.

[2]　Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[J]. arXiv:1506.02640, 2015.

[3]　Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[4]　He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[5]　Girshick R. Fast R-CNN[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision, Santiago, Dec 7-13, 2015. Piscataway: IEEE, 2015: 1440-1448.

[6]　Ren S, He K, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal net-

works[C]//Proceedings of the 28th International Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2015: 91-99.

[7] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, Jul 21-26, 2017. Washington: IEEE Computer Society, 2017: 7263-7271.

[8] Redmon J, Farhadi A. Yolov3: an incremental improvement[J]. arXiv:1804.02767, 2018.

[9] Bochkovskiy A, Wang C Y, Liao H Y M. YOLOv4: optimal speed and accuracy of object detection[J]. arXiv:2004.10934, 2020.

[10] Dai J, Li Y, He K, et al. R-FCN: object detection via region-based fully convolutional networks[C]// Proceedings of the 30th International Conference on Neural Information Processing Systems. Red Hook: Curran Associates Inc., 2016: 379-387.

[11] Wang L, Guo S, Huang W, et al. Places 205-vggnet models for scene recognition[J]. arXiv:1508.01667, 2015.

[12] Sasha T, Almeida D, Lyman K. Resnet in resnet: generalizing residual architectures[J]. arXiv:1603.08029, 2016.

JIANG Xinlan was born in 1976. She received the Ph.D. degree in computer science and technology from Beijing Jiaotong University in 2018. Now she is a lecturer at University of Chinese Academy of Social Sciences. Her research interests include computer vision, machine learning, optical inspection, etc.

蒋欣兰（1976-），女，江苏常州人，2018 年于北京交通大学计算机科学与技术专业获得博士学位，现为中国社会科学院大学讲师，主要研究领域为计算机视觉，机器学习，光学检测等。发表学术论文 10 余篇。