

# 基于改进 Mask RCNN 的电力检修违规操作检测<sup>①</sup>



沈茂东<sup>1</sup>, 周伟<sup>1</sup>, 宋晓东<sup>1</sup>, 裴健<sup>1</sup>, 邓昊<sup>2</sup>, 马超<sup>2</sup>, 房凯<sup>3</sup>

<sup>1</sup>(国网山东省电力公司, 济南 250001)

<sup>2</sup>(山东鲁能软件技术有限公司, 济南 250001)

<sup>3</sup>(中国石油大学(华东) 计算机科学与技术学院, 青岛 266580)

通讯作者: 房凯, E-mail: k2322358928@163.com

**摘要:** 电力检修现场中施工行为的规范关系到工作人员的人身安全, 对电力行业的发展至关重要. 为了从计算机视觉的角度对电力检修工作人员的违规操作行为进行检测, 基于 Mask RCNN 算法设计了一种多任务多分支违规行为检测算法, 综合目标检测、关键点检测与实例分割任务, 并行对目标进行检测并获得目标的边框坐标、关键点与 mask 信息. 实验结果表明, 相对于原来的 Mask RCNN, 该算法在实例分割和关键点检测方面都有了显著的提升, 具有更高的准确性和鲁棒性, 在电力检修违规检测方面满足实际部署的精度要求.

**关键词:** 多分支网络; 深度学习; 行为分析; 目标检测; 违规检测

引用格式: 沈茂东, 周伟, 宋晓东, 裴健, 邓昊, 马超, 房凯. 基于改进 Mask RCNN 的电力检修违规操作检测. 计算机系统应用, 2020, 29(8): 158-164. <http://www.c-s-a.org.cn/1003-3254/7559.html>

## Illegal Operation Detection in Electric Maintenance Based on Improved Mask RCNN

SHEN Mao-Dong<sup>1</sup>, ZHOU Wei<sup>1</sup>, SONG Xiao-Dong<sup>1</sup>, PEI Jian<sup>1</sup>, DENG Hao<sup>2</sup>, MA Chao<sup>2</sup>, FANG Kai<sup>3</sup>

<sup>1</sup>(State Grid Shandong Electric Power Company, Jinan 250001, China)

<sup>2</sup>(Shandong Luneng Software Technology Co.Ltd, Jinan 250001, China)

<sup>3</sup>(College of Computer Science and Technology, China University of Petroleum, Qingdao 266580, China)

**Abstract:** The norm of operation in electric power maintenance is related to the personal safety of the staff, and is very important to the development of electric power industry. In order to detect the illegal operation behavior of power maintenance workers from the perspective of computer vision, a multi-tasking and multi-branch illegal behavior detection algorithm was designed based on the Mask RCNN algorithm. It integrates target detection, key point detection and instance segmentation tasks, and performs parallel target detection. Detect and obtain the frame coordinates, keypoints, and mask information of the target. The experimental result demonstrates that this algorithm has significantly improved the precision in instance segmentation and key point detection, has higher accuracy and robustness compared with Mask RCNN. And it meets the accuracy requirements of actual deployment in power maintenance violation detection.

**Key words:** multi-branch network; deep learning; behavior analysis; object detection; illegal operation detection

## 1 引言

在电力检修行业中施工行为是否规范关系到工作人员的人身安全, 对于电力行业的发展至关重要<sup>[1]</sup>. 人

员行为分析是当前深度学习领域的一个重要的研究方向, 目前在电力行业的检修工作中已经基本实现了视频监控全覆盖, 通过对这些视频进行分析可以获得工

① 基金项目: 国家重点研发计划 (2017ZX05013-002); 山东省自然科学基金 (ZR2019MF049)

Foundation item: National Key Research and Development Program of China (2017ZX05013-002); Natural Science Foundation of Shandong Province (ZR2019MF049)

收稿时间: 2020-01-08; 修改时间: 2020-02-08; 采用时间: 2020-03-11; csa 在线出版时间: 2020-07-29

作人员施工的相关信息. 针对电力检修工作人员的违规操作问题, 本文于从计算机视觉的角度对电力检修现场的违规行为进行检测, 基于 Mask RCNN<sup>[2]</sup> 设计了一种多任务多分支违规行为检测网络, 综合目标检测、关键点检测与实例分割任务, 以获得违规行为检测所需的目标坐标、mask 掩码与关键点信息. 同时为了避免实际应用中由于训练数据不足的过拟合问题, 设计了生成式数据增强与传统几何变换相结合的数据增强<sup>[3]</sup> 策略来扩充数据集并通过交叉验证实验选取了最佳的超参数. 实验结果表明, 相对于原先的 Mask RCNN, 基于改进 Mask RCNN 的电力维修违规检测方法具有更高的准确性和鲁棒性, 满足实际部署的精度要求.

## 2 相关工作

自从 Hinton 等<sup>[4]</sup> 在 2012 年提出深度学习的概念, 深度学习便逐渐取代了传统算法成为目标检测领域的主流方法. 到目前为止, 深度学习在图像处理领域已经取得了令人瞩目的成就. 近年来深度学习的发展为违规行为检测提供了新的方案, 卷积神经网络对于目标的几何变换、形变、光照等因素适应性较强, 有效克服了由于目标外观的多样性带来的识别阻力. 它可根据输入到网络的数据自动地生成相应特征的描述, 具有较高的灵活性和普适性<sup>[4,5]</sup>. 目前图像中的目标检测主要分为单步检测方法(如 SSD, YOLO 和 CornerNet<sup>[6-8]</sup>)和两步检测方法(如 R-CNN、Fast R-CNN、Faster R-CNN<sup>[9-11]</sup>和 Mask R-CNN<sup>[2]</sup>)两种. 单步检测方法直接在图像上经过计算生成检测结果. 两步检测方法先在图像上提取候选区域, 再基于候选区域进行特征提取然后再进行预测. 单步检测算法的主要思想就是直接从输入图像得到目标物体的类别和具体位置, 不再像两步检测算法系列那样产生候选区域. 单步检测方法最显著的优势就是效率, 在检测精度上相对于两步检测方法则稍逊一筹.

语义分割是计算机视觉领域中的一个典型问题, 一般将其视作按照图像的语义来将图像进行分割, 即将图像中的每一个像素进行分类. 实例分割则比语义分割更进一步, 首先在图像中框出不同的实例, 然后再用语义分割的方法在不同的实例框中为每个区域的像素进行标记, 可将其视为目标检测与语义分割的结合体. FCN<sup>[12]</sup> 是图像分割领域的里程碑式的突破, FCN 全名为全卷积网络, 顾名思义, 即不包含全连接层的网络. FCN 将 CNN 中的全连接层用  $1 \times 1$  卷积代替, 可适

应任意尺寸的输入, 并利用反卷积 (DeConvolution) 来实现对特征图的上采样, 最终得到具有较高准确性的分割图像. FCN 实现了图像的端到端分割, 但是仍存在一定的缺陷, 如效率不够高、分割不够精细等. U-Net<sup>[13]</sup> 是最近提出的应用于生物医学图像分割的卷积网络, 与 FCN 网络不同的是, U-Net 的上采样阶段与下采样阶段采用了相同数量层次的卷积操作, 且使用 skip connection 结构将下采样层与上采样层相连, 使得下采样层提取道的特征可以直接传递到上采样层, 这使得 U-Net 网络的像素定位更加准确, 分割精度更高.

深度学习在关键点检测方向上的应用主要涉及人脸关键点检测以及用于姿态估计的人体关键点检测. Openface<sup>[14]</sup> 是当前较为流行的一个人脸关键点检测框架, 基于 Schroff 等在 CVPR2015 上提出的 Facenet<sup>[15]</sup>, 该算法首先使用 Dlib 或者 OpenCV 检测到人的面部, 然后利用仿射变换使眼睛等标志点的位置固定, 具有较高的检测速度. CVPR2017 上提出的 PAF (Part Affinity Fields)<sup>[16]</sup> 以一种自底向上的方式实现了在一整张图像中对其中多人进行关键点检测. 该网络分为两大部分: 一部分根据特征图进行关键点预测; 另一部分获得每个关键点的 PAF. 将这两部分进行联合学习和预测, 将这些关键点进行两两连接, 进而转化为图论问题.

何恺明在 2017 年提出的 Mask RCNN 是当前工业界最为流行的目标检测与分割框架之一, 既能准确检测出图像中的目标, 又能为一个目标生成具有较高质量的 mask. Mask RCNN 实际上是 FasterRCNN 的一个拓展, 相对于 FasterRCNN, MaskRCNN 做了如下改进:

- 1) 用 AoIAlign 代替 RoIPooling, 实现了特征图的像素级对齐;

- 2) 在 Faster RCNN 每个 RoI 类标签和坐标两个输出的基础上添加了第三个 Mask 分支, 在每个 RoI 上利用 FCN 以  $\text{pix}2\text{pix}$  的方式实现 Mask 分割;

- 3) Mask 分割分支易于向关键点检测分支拓展.

但是上述方法均为一个或者两个任务并行检测, 无法同时实现目标检测、实例分割和关键点检测任务, 在对人与物体有交互行为的电力检修违规检测问题上变得不再适用. 因此需要根据电力检修的实际需要, 结合现实中人与物体交互中的违规操作, 对 Mask RCNN 算法进行改进, 设计一种可以并行实现目标检测、实例分割与关键点检测的多任务多分支检测算法.

### 3 基于改进 Mask RCNN 的电力检修违规行为检测

针对电力检修现场施工中的违规行为检测问题, 本文对 Mask RCNN 进行改进提出了一种多任务多分支的违规行为检测网络, 该方法在 Mask RCNN 的基础上将目标检测、关键点检测与实例分割在一个网络中整合, 实现了一种端到端的多任务检测算法; 同时, 将实例分割从目标检测分支中分离出来, 消除由目标检测边界框偏差带来的实例分割偏差; 对于分类误差、边框误差、实例分割误差和关键点检测误差, 设计了新的损失函数; 相对应用场景, 所得到的数据集偏少, 为避免实际应用中的过拟合问题, 设计了生成式数据增强与传统几何变换相结合的数据增强<sup>[3]</sup>策略来扩充数据集.

#### 3.1 数据增强

由于可以获得的数据集样本较少, 采用传统图像几何变换数据增强和利用 GAN<sup>[17]</sup> 进行生成式数据增强相结合的方法对数据集进行扩充. 对于传统的图像几何变换方法, 采用随机裁剪、水平翻转、图像倾斜、添加噪声、图像缩放和 mixup<sup>[18]</sup> 的方式对数据集进行了扩充, 图 1 为添加高斯噪声的图像.



图 1 添加高斯噪声图像示例

在用 GAN 对数据集进行扩充时, 采用 DCGAN 来生成新的图像<sup>[19]</sup>. DCGAN 是比 GAN 拥有更多隐藏层的一种特殊 GAN 结构, 它与 GAN 在结构上的不同在于: 一是 DCGAN 中没有池化层 (pooling 层), 并且在生成器中使用转置卷积进行上采样而在判别器中使用微步幅卷积; 二是 DCGAN 去掉了全连接层从而使网络

变成了全卷积网络, 这样的做法提高了输出图像的质量并且提高了模型的收敛速度.

如图 2 是本文采用的基于 DCGAN 的图像生成模型结构示意图. 其中生成器有 15 层, 判别器有 16 层, 在卷积层后使用了批归一化处理, 在生成器和判别器中均采用 Leaky-ReLU 作为激活函数.

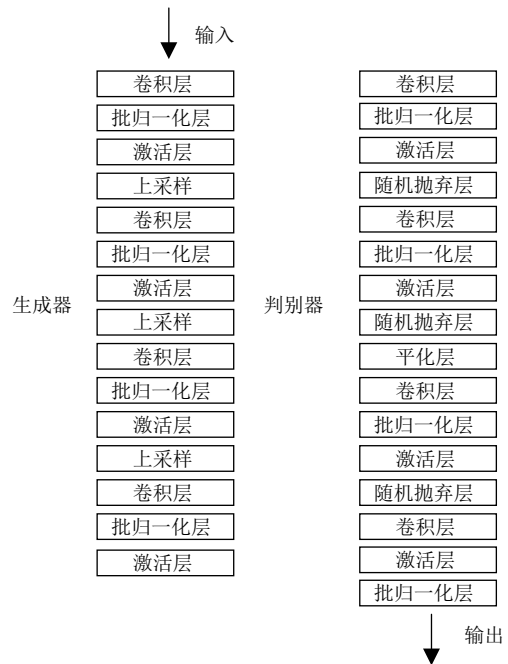


图 2 图像生成模型结构示意图

考虑计算机显存的限制, 在训练时统一将图像在 800×600 的尺寸上进行训练, 每 16 张图像为一个批次, 在生成器和判别器中均采用 Adam 作为优化器. 训练过程中生成的图像样本如图 3 所示.



图 3 生成图像示例

### 3.2 改进的 Mask RCNN 网络结构

经过改进的 Mask RCNN 实现了多个计算机视觉任务的并行处理, 该网络主要包括两个阶段: 第 1 阶段主要利用区域建议网络 (RPN) 生成可能存在目标的候选区域边界框; 在第 2 阶段中对第 1 阶段生成的每个候选区域利用 RoIAlign 进行对齐, 然后分别完成分类、边框回归、Mask 预测及关键点预测。

如图 4 所示, 输入图像经过缩放以后输入特征提取骨架网络, 生成更具有更高阶特征的特征图, 以实现后面多个分支与 RPN 的特征共享。在此选择的是 ResNet101,

同时添加 FPN 以解决目标的多尺寸问题。多个尺寸的特征图经过 RPN 生成多个无类别的候选区域, 并利用非极大值抑制 (NMS) 消除大量冗余候选区域, 这些候选区域经过 RoIAlign 在每一个分支上归一化成同一尺寸的特征图。在 RPN 部分, 对于每一个三通道的图像中的每一个位置, 都会生成 9 个候选窗口: 3 个纵横比  $\{1:1, 1:2, 2:1\} \times 3$  种尺寸  $\{128, 256, 512\}$ 。考虑到实际应用中瓷瓶、人体等目标的大小和比例, 对上述纵横比指标进行了相应的调整, 将原来的纵横比调整为  $\{1:2, 2:1, 1:3, 3:1\}$ , 以此实现前景与背景的精准区分。

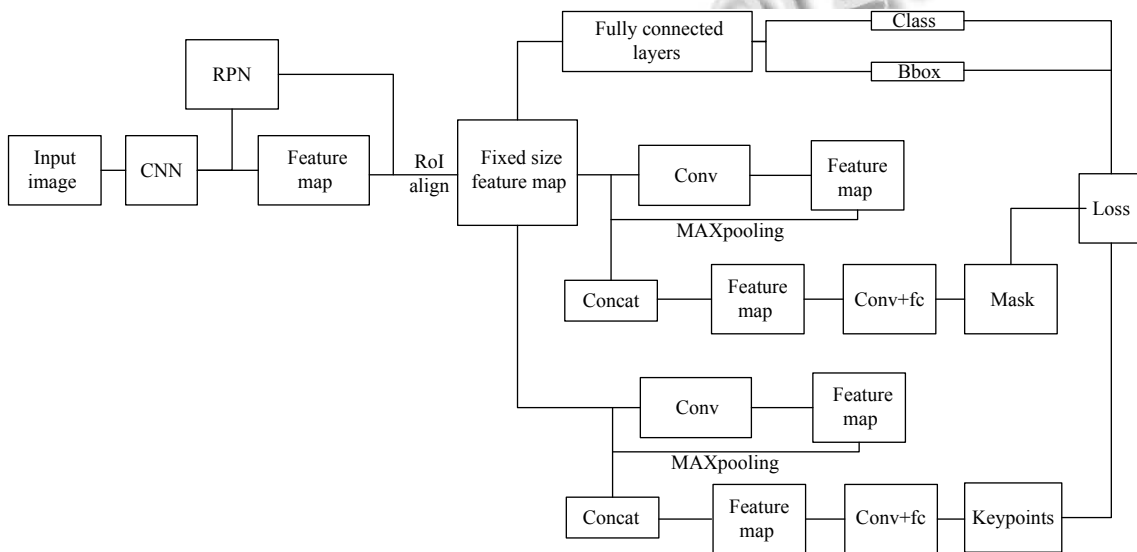


图 4 改进的 Mask RCNN 网络结构图

对于分类和边框回归, 经过归一化的特征图通过全连接层预测每个 RoI 的所属类别, 并对其坐标进行回归。

对于 Mask 预测, 与 Mask RCNN 不同的是, 在改进的 Mask RCNN 网络中 Mask 预测独立进行, 而不是依赖于目标的边框坐标, 以此消除由坐标偏差导致的实例分割误差。Mask 预测分支将 Mask 预测分为两个部分, 类别预测和 Mask 回归, 由于在 Mask 类别预测上要与第一个分支一致, 在此可以直接引用第一个分支的类别信息而专注于 Mask 的回归。因此需要对 Mask 的质量设计一个新的评估指标 MaskIoU, 即真实 Mask 与预测 Mask 的交并比。

最后, 与 Mask 分支类似, 关键点检测分支在特征图上利用线性回归得到对应的预测位置。

### 3.3 损失函数设计

对于各个分支中的各个任务而言, 其重要性是相

同的, 但是在多任务学习中, 往往不同任务的学习难易程度不同, 若采用相同的损失权重, 会导致训练难以收敛。如式 (1), 针对各个任务, 分别设计独立的损失函数, 加权相加作为最终损失, 即损失=分类误差+坐标误差+分割误差+关键点误差,  $\delta$ 、 $\varepsilon$ 、 $\theta$  均为加权系数。

$$Loss = L_{cls} + \delta L_{bbox} + \varepsilon L_{mask} + \theta L_{kps} \quad (1)$$

$L_{cls}$ 、 $L_{bbox}$  即全连接层预测出的每个 RoI 的所属类别及其矩形框坐标值误差。类别损失函数如式 (2) 所示, 其中  $p_i$  表示预测为目标概率,  $p_i^*$  用 0 和 1 表示是否为真实目标。

$$L_{cls} = \sum_i \{-\log [p_i^* p_i + (1 - p_i^*) (1 - p_i)]\} \quad (2)$$

目标坐标的损失函数如式 (3) 所示, 其中  $(x_i, y_i)$  表示第  $i$  预测结果的中心点坐标,  $(x_i^*, y_i^*)$  为真实的中心点坐标。相应地,  $(w_i, h_i)$  和  $(w_i^*, h_i^*)$  分别代表预测目标和真

实目标的宽高.

$$L_{\text{bbox}} = \sum_i p_i^* \begin{bmatrix} x_i \log x_i^* + \\ y_i \log y_i^* + \\ \sqrt{w_i} \log \sqrt{w_i^*} + \\ \sqrt{h_i} \log \sqrt{h_i^*} + \\ (1-x_i)(1-\log x_i^*) + \\ (1-y_i)(1-\log y_i^*) + \\ (1-\sqrt{w_i})(1-\log \sqrt{w_i^*}) + \\ (1-\sqrt{h_i})(1-\log \sqrt{h_i^*}) \end{bmatrix} \quad (3)$$

Mask 损失函数的设计主要侧重为: 将 Mask 划分至正确的类别, 并对前景对象类别的 Mask 进行回归. 但是在一个损失函数中难以同时实现两个功能, 并且 Mask 类别与目标分类分支的类别是一致的. 为了简化 Mask 类别损失可以直接取分类损失  $L_{\text{cls}}$ , 而边界损失则可以用真实 Mask 与经过某一阈值二值化的预测 Mask 的交并比表示. 此时 Mask 损失函数可如下所示:

$$L_{\text{mask}} = L_{\text{cls}} * \text{MaskIoU} \quad (4)$$

其中,  $\text{MaskIoU}$  有如下定义:

$$\text{MaskIoU} = \frac{\text{mask}_{\text{gt}} \cap \text{mask}_{\text{pre}}}{\text{mask}_{\text{gt}} \cup \text{mask}_{\text{pre}}} \quad (5)$$

而关键点检测的损失函数则较为简单, 只需计算真实值与预测值之间的  $L_1$  损失, 为了使损失函数更加平滑, 在此使用时  $\text{Smooth}_{L_1}$  作为关键点检测的损失函数.

$$L_{\text{kps}} = \text{Smooth}_{L_1}(kp_i - kp_i^*) \quad (6)$$

其中,

$$\text{Smooth}_{L_1}(x) = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & |x| > 1 \end{cases} \quad (7)$$

#### 4 实验与分析

本次实验以检测人员在瓷瓶上挂安全带为例, 训练集图像共有 3000 张, 通过数据增强将训练集扩充至 20 000 张. 在模型训练与测试时采取的是 10 折交叉验证的方式, 即将数据均分为 10 份, 每次训练集取其中 9 份, 留 1 份作为测试集, 最终结果为 10 次平均值.

实验显卡配置 NVIDIA GeForce GTX TITAN X, 处理器为 Intel i7.

在本实验中, 以目标检测任务为例. 首先, 我们要先对所有的检测结果排序, 得分越高的排序越靠前, 然后依次判断检测是否成功. 将排序后的所有结果定义

为 DTs (预测框), 所有同类别的真实目标定义为 GTs (真实框). 先依序遍历一遍 DTs 中的所有 DT, 每个 DT 和全部 GT 都计算一个 IoU (bbox 的交并比), 作为检测任务性能的评价指标, 如果最大的 IOU 超过了给定的阈值, 那么视为检测成功, 算作 TP (True Positive); 如果该 DT 与所有 GT 的 IOU 都没超过阈值, 自然就是 FP (False Positive). 同理, 实例分割及关键点检测任务性能的评价与目标检测类似.

实验过程中, 对于目标检测、实例分割、关键点检测分别有各自的评价指标, 并分别设置了相对应的阈值, 具体指标如表 1 所示.

表 1 评价指标

任务点	评价指标	评价度量阈值
目标检测	预测值与真实值 bbox 的交并比	0.5
实例分割	预测 Mask 与真实 Mask 交并比	0.5
关键点检测	预测点与真实点的欧氏距离	10 pixel

#### 4.1 改进的 MaskRCNN 效果测试

如图 5 为 Mask RCNN 检测算法测试效果图, 从图中可以看出, 在此场景下存在人员漏检的情况.



图 5 Mask RCNN 检测算法测试效果图

图 6 为改进的 Mask RCNN 多任务检测算法测试效果图. 从图中可以看出, 使用改进后的 Mask RCNN 多任务检测算法可以更为准确地检测出相关目标信息, 综合了目标检测、关键点检测与实例分割任务, 并且可以通过施加额外的判断逻辑对瓷瓶上挂安全带这一违规行为准确检测. 其中红框为瓷瓶上安全带的检测结果.

果, 蓝框则表示存在人员在瓷瓶上挂安全带的违规行为, 表 2 为检测效果对比。



图 6 改进的 Mask RCNN 多任务检测算法测试效果图

表 2 改进的 Mask RCNN 与 Mask RCNN 检测效果对比

模型	目标检测	实例分割	关键点检测
Mask RCNN	92.3	89.2	89.6
改进的MaskRCNN	93.2	92.3	91.1

#### 4.2 测试超参数选择

在模型测试时, 发现电力检修现场中由于遮挡或者各种检修设备的干扰, 存在一定的误检或者将一个目标检测为两个的现象。为了解决此类问题, 可以采用 NMS 的方法对检测到的冗余目标进行消除, 不同的 NMS 数值与准确率 (*Precision*)、召回率 (*Recall*) 的关系如表 3 所示中准确率和召回率有如下定义, 其中 TP 代表检测正确的结果数量, FN 表示漏检的数量, FP 表示误检测结果数量。

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

从表中可以看出, 随着 NMS 的递增, 误检逐步增加, 而漏检逐步递减。原因在于 NMS 越小, 两个检测框交并比在较小的时候就视为同一个目标, 这样就会尽可能地消除冗余框。经过对准确率和召回率进行权衡, 最终在实际应用中选择 0.4 作为 NMS 值。

#### 4.3 数据增强

为了测试该方法所采用的数据增强技术对准确率

的影响, 对无数据增强、传统的图像几何变换、GAN、GAN 与图形变换相结合的方式进行了对照实验, 其中原始图像 3000 张, 传统的图像几何变换生成新的图像 7000 张, GAN 生成新图像 10 000 张。实验结果如表 4 所示, 可以看出, GAN 与传统图像几何变换相结合的方式可以极大地提高模型的准确率。

表 3 NMS 对测试结果影响 (%)

NMS	准确率	召回率
0.1	96.22	98.85
0.2	98.13	98.67
0.3	98.40	98.53
0.4	98.42	95.20
0.5	98.45	93.54
0.6	98.90	91.13
0.7	99.10	84.79
0.8	99.30	84.62
0.9	99.37	75.10

表 4 对照实验 (%)

内容	准确率	召回率
原图	78.1	71.5
几何变换	85.7	81.3
GAN	93.6	90.5
GAN+几何变换	95.2	98.4

## 5 结论与展望

针对电力检修工作中的违规行为检测问题, 本文对 Mask RCNN 算法进行了改进, 主要贡献包括实现了目标检测、实例分割、关键点检测的多任务端到端检测; 将实例分割分支与目标检测分支相分离, 避免由于目标检测分支误差给实例分割带来的负面影响; 针对数据集不足的问题采用多种方式进行数据扩充; 结合实际应用情况, 利用交叉验证选取了最优的 NMS 超参数。实验结果表明, 相对于原来的 Mask RCNN, 该算法在准确性上有了显著的提升, 具有更高的准确性和鲁棒性, 在电力检修违规检测方面满足实际部署的精度要求。在未来还将考虑将人与物体的交互赋予可学习的判断逻辑, 实现端到端的违规操作检测。

#### 参考文献

- 刘洋. 略论电力施工和检修的安全控制策略. 河南科技, 2015, (22): 193. [doi: 10.3969/j.issn.1003-5168.2015.22.155]

- 2 He KM, Gkioxari G, Dollár P, *et al.* Mask R-CNN. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy. 2017. 2980–2988.
- 3 Lamburt L, Koyfman L. Data enhancement techniques: USA, 6397228. 2002-05-28.
- 4 Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Proceedings of the 25th International Conference on Neural Information Processing Systems. Lake Tahoe, NV, USA. 2012. 1097–1105.
- 5 LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436–444. [doi: [10.1038/nature14539](https://doi.org/10.1038/nature14539)]
- 6 Liu W, Anguelov D, Erhan D, *et al.* SSD: Single shot multibox detector. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands. 2016. 21–37.
- 7 Redmon J, Divvala S, Girshick R, *et al.* You only look once: Unified, real-time object detection. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 779–788.
- 8 Law H, Deng J. Cornernet: Detecting objects as paired keypoints. Proceedings of the 15th European Conference on Computer Vision. Munich, Germany. 2018. 765–781.
- 9 Girshick R, Donahue J, Darrell T, *et al.* Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, USA. 2014. 580–587.
- 10 Girshick R. Fast R-CNN. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile. 2015. 1440–1448.
- 11 Ren SQ, He KM, Girshick R, *et al.* Faster R-CNN: Towards real-time object detection with region proposal networks. Proceedings of Advances in Neural Information Processing Systems. Montreal, QB, Canada. 2015. 91–99.
- 12 Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 3431–3440.
- 13 Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany. 2015. 234–241.
- 14 Amos B, Ludwiczuk B, Satyanarayanan M. OpenFace: A general-purpose face recognition library with mobile applications. Pittsburgh: Carnegie Mellon University, 2016. 6.
- 15 Schroff F, Kalenichenko D, Philbin J. Facenet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA. 2015. 815–823.
- 16 Cao Z, Simon T, Wei SE, *et al.* Realtime multi-person 2D pose estimation using part affinity fields. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA. 2017. 1302–1310.
- 17 Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, UK. 2014. 2672–2680.
- 18 Zhang HY, Cisse M, Dauphin YN, *et al.* Mixup: Beyond empirical risk minimization. arXiv: 1710.09412, 2017.
- 19 Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv: 1511.06434, 2015.