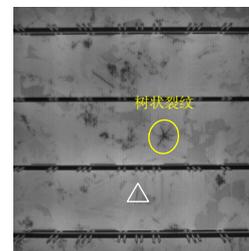




DOI: 10.12086/oe.2021.200112

融合注意力的多尺度 Faster RCNN 的裂纹检测

陈海永^{1*}, 赵鹏¹, 闫皓炜²¹河北工业大学人工智能与数据科学学院, 天津 300000;²天津航天中为数据系统科技有限公司, 天津 300000

摘要: 电致发光(Electroluminescence, EL)下的光伏电池 EL 图像背景表现为复杂的非均匀纹理特征, 且存在与裂纹相似的晶粒伪缺陷, 同时裂纹表现为形状多样的多尺度特征, 以上难点为检测任务带来了极大的挑战。因此, 本文提出融合注意力的多尺度 Faster-RCNN 模型, 一方面, 采用改进的特征金字塔网络获取多尺度的高级语义特征图, 以此来提高网络对多尺度裂纹缺陷的特征表达能力。另一方面, 采用改进的注意力区域推荐网络 A-RPN, 提高模型对裂纹缺陷的关注并抑制复杂背景及晶粒伪缺陷的特征。同时, 在 RPN 网络训练过程中, 采用损失函数 Focal loss, 以此来降低训练过程中简单样本所占比重, 使其更加关注难以区分的样本。实验结果表明, 改进的算法使得 EL 图像裂纹缺陷检测的准确率提高, 达到接近 95%。

关键词: 多尺度特征提取; 注意力模块; Focal loss 函数

中图分类号: TP391.41

文献标志码: A

陈海永, 赵鹏, 闫皓炜. 融合注意力的多尺度 Faster RCNN 的裂纹检测[J]. 光电工程, 2021, 48(1): 200112

Chen H Y, Zhao P, Yan H W. Crack detection based on multi-scale Faster RCNN with attention[J]. *Opto-Electron Eng*, 2021, 48(1): 200112

Crack detection based on multi-scale Faster RCNN with attention

Chen Haiyong^{1*}, Zhao Peng¹, Yan Haowei²¹School of Artificial Intelligence, Hebei University of Technology, Tianjin 300000, China;²Tianjin Aerospace Zhongwei Data System Technology Co., Ltd, Tianjin 300000, China

Abstract: The background of the EL image of a photovoltaic cell under electroluminescence (EL) presents complex non-uniform texture features, and there are grain pseudo-defects similar to cracks. At the same time, the cracks appear as multi-scale features with various shapes. The above mentioned difficulties have presented great challenges for the detection task. Therefore, this paper proposes a multi-scale Faster-RCNN model that integrates attention. On the one hand, an improved feature pyramid network is used to obtain multi-scale advanced semantic feature maps to improve the network's feature expression ability of multi-scale crack defects. On the other hand, an improved attention region proposal network A-RPN is adopted to increase the model's attention to crack defects and suppress the characteristics of complex background and grain pseudo-defects. At the same time, in the RPN network training process, a loss function Focal loss is used to reduce the proportion of simple samples in the training

收稿日期: 2020-04-02; 收到修改稿日期: 2020-06-15

基金项目: 国家自然科学基金资助项目(61873315)

作者简介: 陈海永(1980-), 男, 博士, 教授, 主要从事计算机视觉的研究。E-mail: haiyong.chen@hebut.edu.cn

版权所有©2021 中国科学院光电技术研究所

process, so that the model pays more attention to the samples that are difficult to distinguish. Experimental results show that this algorithm improves the accuracy of crack defect detection in EL images, reaching nearly 95%.

Keywords: multi-scale feature extraction; attention module; focal loss function

1 引言

由于太阳能清洁、安全、无害、充足的特点,光伏发电已成为最受欢迎的可再生能源发电方式之一。光伏电池作为光伏发电的载体,在其生产、部件组装以及运输过程中,因晶体结构的脆弱性或者操作不当等不稳定因素的影响,不可避免地会产生裂纹缺陷。裂纹缺陷的存在将降低电池片的发电效率、缩短使用寿命、导致光伏组件运行故障,甚至影响光伏发电系统的安全性。因此,实现光伏电池的裂纹缺陷检测具有十分重要的现实意义。

Anwar 等人^[1]利用图像分割和形状分析技术提取裂纹的形状特征作为支持向量机(support vector machine, SVM)分类器的输入,实现了电致发光(electroluminescence, EL)图像的微裂纹缺陷检测。Chen 等人^[2]为了获得表达能力更强的缺陷特征,对局部二值模式(local binary patterns, LBP)^[3]进行改进,提出了一种新颖的特征描述符,称为中心像素信息中心对称局部二值模式(CPICS-LBP),并结合 SVM 分类器完成了低分辨率 128×128 像素的 EL 图像块的裂纹缺陷识别。Tsai 等人^[4]提出一种各向异性扩散方案,在生成的扩散图像中进行二进制阈值处理后,通过形态学运算以检测具有低灰度和高梯度特性的微小裂纹缺陷。上述人工定义缺陷特征的表达能力有限,对于 EL 图像中形状多样的多尺度裂纹缺陷而言,算法的鲁棒性和泛化能力不够好。

随着计算机软硬件的发展,深度学习算法被逐渐应用于工业缺陷检测领域,并获取了较为出色的性能。Young-Jin 等人^[5]采用卷积神经网络(convolutional neural networks, CNN)与滑动窗口技术相结合,实现了混凝土裂纹的检测,但模型无法获得输入图像的全局特征信息且难以找到最佳的滑动窗口尺寸。Lin 等人^[6]利用传统的卷积神经网络生成热图的方式实现了背景均匀的 LED 芯片图像的缺陷识别以及定位。如今多种深度学习目标检测算法被提出,如一阶段网络 Yolo, CenterNet^[7], RetinaNet^[8]; 二阶段网络 Fast RCNN^[9]、Faster RCNN^[10]等。相比于一阶段目标检测网络,二阶段将特征提取与区域推荐网络相结合,使得网络实现了端到端检测,检测精度也相对更高。Cha

等人^[11]基于 Faster RCNN 算法开发了实时损伤检测的框架,有效地检测五种类别的表面缺陷。为适应目标尺度变化,高琳等人^[12]结合特征金字塔网络提出了一种融合多尺度上下文卷积的车辆目标检测算法。为了增强低层特征与最顶层特征之间的特征联系, Liu 等人^[13]在特征金字塔结构的基础上添加自下而上的路径,提出了实例分割网络 PANet。为了使模型对感兴趣区域更加关注,注意力机制被广泛研究。人类视觉系统在分析场景时,并不会立即针对整个图像进行处理,而是会选择性地关注于突出的部分,即注意力在人类感知中起到了重要作用^[14]。在 CNN 中,注意力模块起着与人类的注意力感知类似的作用,其通过赋予特征图中的缺陷区域更高的权重,以实现强调目标对象特征的目的,抑制图像背景的噪声干扰。

光伏电池 EL 图像是电池片采用电致发光^[15]技术,由近红外 CCD 相机在 EL 成像采集系统(图 1)中捕获的,图像像素大小为 1024×1024。

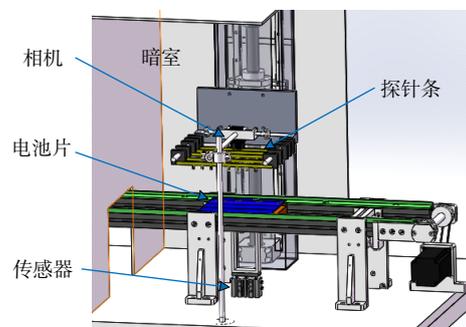


图 1 EL 成像采集系统

Fig. 1 EL imaging acquisition system

对于光伏电池 EL 图像,其检测难点主要有:

1) 任意样本之间的非均匀纹理复杂背景均不相同,具有较高转换效率的无缺陷区域表现为较亮的特点,裂纹缺陷和晶粒表现为更暗。现有的目标检测算法不能很好地抑制复杂背景对裂纹检测的干扰。2) EL 图像背景包含主栅线、副栅线、晶粒等结构,由于其特殊的生产工艺,晶粒具有大小不同、方向不一且随机分布的特点。更重要的是,部分晶粒伪缺陷的结构与裂纹缺陷具有极高的相似性,这使得检测中极易出

现误检。3) 第三, 裂纹缺陷的尺寸、形状的多样性。尺寸多样性表现为裂纹大小多样、纵横比多样; 形状多样性具体表现为横向、纵向和拓扑结构的复杂树状裂纹, 见图 2。

针对以上难点, 本文的主要贡献有: 1) 采用改进的特征金字塔网络获取高分辨率、多尺度的高级语义特征图, 以此来提高网络在复杂背景下对多尺度裂纹缺陷的特征表达能力; 2) 采用改进的注意力区域推荐网络(attention region proposal network, A-RPN), 关注裂纹特征并抑制复杂背景及其中的晶粒伪缺陷的特征; 3) 在 RPN 训练过程中采用分类损失函数 Focal loss, 以此来降低训练过程中简单样本所占比重, 使模型更加关注难以区分的样本, 进一步提升检测效果。总的来说, 提出的融合注意力的多尺度 Faster-RCNN

模型, 在解决非均匀纹理复杂背景的 EL 图像的裂纹缺陷检测任务中, 表现出更优的检测性能。

2 EL 图像裂纹检测算法

Faster-RCNN 包括特征提取网络、区域推荐网络 RPN、感兴趣区域池化 ROI pooling, 以及分类回归模块四个部分。基于此, 本文提出的融合注意力的多尺度 Faster RCNN 模型, 首先通过残差网络 ResNet50 与改进的路径聚合特征金字塔网络(path aggregation feature pyramid network, PA-FPN)相结合, 将提取的多尺度特征图输入融合卷积注意力模块的 A-RPN 网络与 ROI pooling 中, 进一步完成裂纹的检测, 模型结构见图 3。

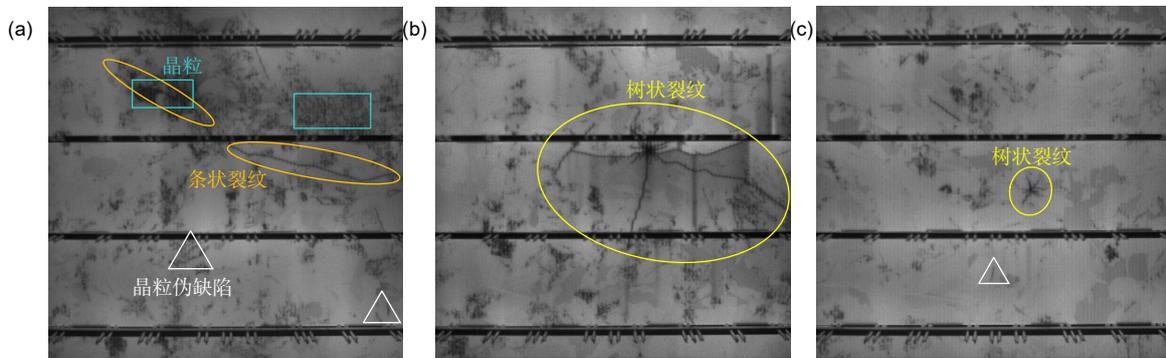


图 2 非均匀纹理随机背景的 EL 图像。

矩形框标记为晶粒, 三角框标记为与裂纹高度相似的晶粒伪缺陷, 椭圆标记为裂纹

Fig. 2 EL image of a random background with a non-uniform texture. The rectangular frame is the grain, the triangular frame marks the pseudo-defects of the grain that are highly similar to the crack, and the ellipse marks the crack

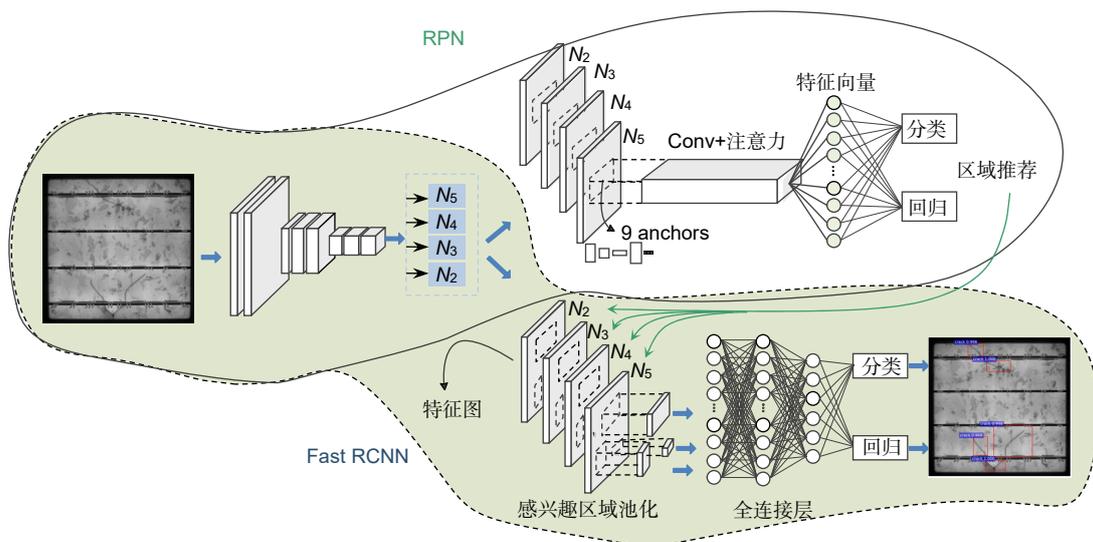


图 3 融合注意力的多尺度 Faster-RCNN 模型

Fig. 3 Multi-scale Faster-RCNN model with attention

2.1 多尺度特征提取网络

本文工作是检测光伏电池 EL 图像中的裂纹缺陷，裂纹表现为尺度变化较大、形状各异的特点。现有的 Faster RCNN 直接将特征提取网络最后一层输出的特征用作后续分类回归，由于浅层网络包含的特征信息易被丢失，会导致小目标裂纹特征无法充分提取，出现漏检问题。本文在特征金字塔网络 FPN 的基础上提出采用改进的路径聚合特征金字塔网络(path aggregation feature pyramid network, PA-FPN)与残差网络 ResNet50 相结合，提升模型对多尺度裂纹缺陷的特征表达能力。

路径聚合特征金字塔网络 PA-FPN 在特征金字塔网络 FPN 的基础上，考虑到浅层网络信息对小目标检测的重要性，添加了自下而上的路径聚合来保留浅层特征。如图 4 所示，实线箭头表示 FPN 的自下而上的前馈计算生成特征 $\{C_2, C_3, C_4, C_5\}$ 。浅层特征经过几十层之后传递到最顶层，使得浅层特征信息丢失较多；FPN 添加了自上而下的计算和横向连接来构建特征金字塔以此获得了多尺度的包含丰富语义信息的特征图 $\{P_2, P_3, P_4, P_5\}$ 。为了更好地保留图像的浅层特征信息，本文添加了自下而上的路径聚合，见图 4 中虚线箭头所示，以不到十层的特征聚合更好地保留了浅层特征信息，得到 $\{N_2, N_3, N_4, N_5\}$ 用于后续的预测目标的分类和回归，其中 N_2 与 P_2 相同。PA-FPN 会提升模型对 EL 图像中的不同尺度裂纹缺陷检测的适应性，尤其是提高小尺寸裂纹目标的检测结果。

2.2 融合注意力的 RPN 网络

2.2.1 Anchor 设置

区域推荐网络 RPN 是 Faster RCNN 的重大改进，RPN 通过滑动窗口为其输入特征图的每一个点配置 k 个锚框 anchor，进一步通过分类器及 bounding box

regression 回归原理完成对目标区域的推荐。设置 anchor 时，在给定基础锚框尺寸的基础上，不仅需要给定不同尺度的 scale 参数，还需要考虑目标的纵横比 ratio 参数。

为了生成更加接近数据集中实际目标缺陷尺寸的锚框，提高算法在缺陷检测过程中对位置回归的速度及准确性，采用 k-means 聚类对 anchor 大小的设置进行指导。针对特定数据集中的目标缺陷尺寸进行聚类，即计算手动标记的 ground-truth 与聚类中心的交并比 (IOU)，以 1-IOU 距离度量进行聚类，指导 RPN 生成与实际缺陷的形状更加匹配的 anchor，进一步完成缺陷的定位与检测。其实现过程如下：

1) 给定 m 个聚类中心点 (W_i, H_i) , $i \in \{1, 2, \dots, m\}$ ，其中， W_i, H_i 表示 anchor 的宽和高，由于其位置并不固定，因此不用给定中心点坐标 (x, y) 。

2) 数据集标注过程中会生成的包含目标位置的文件，其中有 (x_j, y_j, w_j, h_j) , $j \in \{1, 2, \dots, N\}$ ，即相对于原图中 ground-truth 的坐标，其中 (x_j, y_j) 为框的中心点坐标， (w_j, h_j) 为宽和高， N 为标注框的数量。

3) 计算数据集中的标注框与 m 个聚类中心点的距离 d ，其中：

$$d = 1 - \text{IOU}[(x_j, y_j, w_j, h_j), (x_j, y_j, W_i, H_i)],$$

并根据 d 的大小将标注框分到与其最近的 (W_i, H_i) 点。

4) 直到所有标注框分配完毕，对每一个簇计算最终的聚类中心点：

$$W'_i = \frac{1}{N_i} \sum w_i, \quad H'_i = \frac{1}{N_i} \sum h_i,$$

其中： N_i 表示第 i 个簇中包含标注框的个数，整个公式即求簇中所有标注框的平均宽和高的值。

5) 重复 3)、4) 中步骤，直到 (W'_i, H'_i) 变化很小。

针对光伏电池 EL 数据集裂纹缺陷，通过 k-means

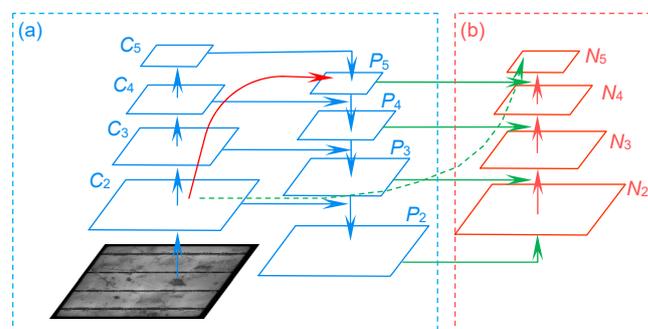


图 4 路径聚合特征金字塔 PA-FPN
Fig. 4 Path aggregation feature pyramid PA-FPN

聚类找到了 anchor 较合适的纵横比 ratio，分别为 {0.25, 0.5, 1, 1.66, 2.1, 3.3}，提高了 RPN 生成的先验框的稳定性。实验证明，增加 anchor 的数量，训练增加的时间并不明显。

2.2.2 注意力模块

为了提高模型对裂纹缺陷特征的关注并抑制复杂背景及晶粒伪缺陷的特征，采用改进的注意力区域推荐网络 A-RPN，引导区域推荐网络 RPN 在光伏电池 EL 图像的复杂背景干扰下更准确地预测裂纹缺陷的位置。如图 5 所示，特征提取网络将输出的特征输入到 A-RPN 网络中，首先对输入特征图进行 3×3 卷积之后得到特征 F 。之后，特征图沿着卷积注意力模块 (convolutional block attention module, CBAM) 获得更加精细的特征 F' ，以便 A-RPN 进行后续更加精确的目标分类与回归。

卷积注意力模块 CBAM 由通道注意力与空间注意力两个互补的模块相连接组成，可以抑制复杂背景的特征同时突出缺陷的特征，并集中于复杂背景下光伏电池 EL 图像中裂纹的空间位置。其中，通道注意力着重于目标“是什么”，通过将包含更多缺陷信息的通道赋予更大的权重，将包含更多背景信息的通道赋予较小的权重，从而来选择包含有用缺陷特征信息的通道。空间注意力会告诉网络缺陷“在哪里”，帮助网络定位缺陷在特征图中的位置。提出注意力区域推荐网络 A-RPN 可以在复杂背景的干扰下生成更加准确

的缺陷推荐区域，进一步提升模型的目标检测效果。

如图 5，假设输入特征图为 $F \in R^{C \times H \times W}$ ，首先经过通道注意力模块^[16]获得注意力图 $M_c \in R^{C \times 1 \times 1}$ ，将 M_c 对输入特征 F 的每一个通道进行加权，而后获得精细化后的特征 F' 。再经过空间注意力模块获得注意力图 $M_s \in R^{1 \times H \times W}$ ，将 $M_s(F')$ 与特征 F' 相乘得到最终输出的 F'' ，表达如下：

$$F' = M_c(F) \otimes F, \quad (1)$$

$$F'' = M_s(F') \otimes F', \quad (2)$$

其中： \otimes 代表矩阵的对应元素相乘， C 表示输入特征的通道数， H 和 W 表示特征图的高度和宽度。

对于关注“是什么”的通道注意力模块，首先对特征图进行平均池化 (avgpooling，在公式中用 P_{avg} 表示) 和最大池化 (maxpooling，在公式中用 P_{max} 表示)，提取两种特征描述 F_{avg}^c 和 F_{max}^c ，其次，将这两个不同的特征分别送入中间的共享网络层，在共享网络应用于 F_{avg}^c 和 F_{max}^c 特征之后，将得到的两个特征的对应元素相加，再经过 sigmoid 激活函数得到通道注意力图 $M_c \in R^{C \times 1 \times 1}$ 。计算公式如下：

$$M_c(F) = \sigma(P_{MLP}(P_{avg}(F)) + P_{MLP}(P_{max}(F))) \\ = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c))) \quad (3)$$

其中：共享网络层表示一个多层感知机 (multi-layer perceptron, MLP，在公式中用 P_{MLP} 表示)，包含两层的神经网络。 σ 代表 sigmoid 激活函数， W_0 和 W_1 代表的是多层感知机模型的两层参数，且 W_0 和 W_1 之间的

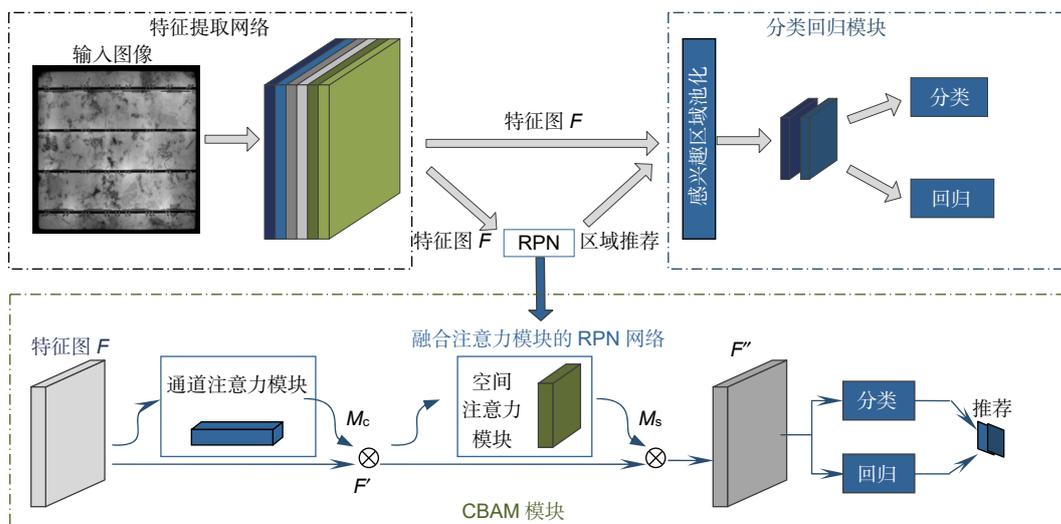


图 5 融合注意力 CBAM 的检测模型
Fig. 5 Detection model with integrated CBAM

特征采用 ReLU 作为激活函数进行处理。最后，将 $M_c(F)$ 与其输入特征相乘得到经过通道注意力调整的精特征图 F' 。

空间注意力模块更加关注“在哪里”，即输入特征图上缺陷的空间位置。空间注意力的输入是通道注意力模块的输出 F' ，通过平均池化和最大池化获得特征图 $F_{avg}^s \in R^{1 \times H \times W}$ 和 $F_{max}^s \in R^{1 \times H \times W}$ ，再将特征图合并后使用 7×7 的卷积核、经过 sigmoid 激活函数得到新的空间注意力图 M_s 。总的来说，空间注意力的计算公式为：

$$M_s(F) = \sigma(f^{7 \times 7}([(P_{avg}(F); P_{max}(F))])) = \sigma(f^{7 \times 7}([F_{avg}^s; F_{max}^s])) \quad (4)$$

其中： σ 代表 sigmoid 激活函数， $f^{7 \times 7}$ 代表卷积核大小为 7×7 的卷积操作。最后，将 $M_s(F')$ 与其输入特征相乘得到最终的特征图 F'' 。

2.2.3 损失函数 Focal loss

在 RPN 网络中，训练过程通过采用交叉熵损失函数来优化参数。对于二分类交叉熵损失函数：

$$L = -y \log y' - (1 - y) \log(1 - y') = \begin{cases} -\log y', & y = 1 \\ -\log(1 - y'), & y = 0 \end{cases} \quad (5)$$

其中： y' 是网络输出某一类别的概率值，取值在 0 到 1 之间。对于正样本而言，输出概率值的大小决定损失函数的大小，概率值越大代表预测越接近真实值，损失函数 L 的值也越小。对于负样本，其包含目标的概率值越小则损失函数的值越小。研究表明^[8]，前景与背景样本比例失衡会降低网络训练效率，且简单样本会主导梯度更新方向。对于光伏电池 EL 图像，复杂背景中包含与裂纹结构高度相似的晶粒伪缺陷，且相比于背景，窄而长的裂纹缺陷所占像素极少。因此，采用损失函数 Focal loss 替代 RPN 网络的分类损失函数，进一步抑制非均匀纹理复杂背景的干扰。

在式(5)中添加 γ 参数后，式(5)变为

$$L_{fl} = \begin{cases} -\alpha(1 - y')^\gamma \log y', & y = 1 \\ -\alpha(1 - \alpha)y'^\gamma \log(1 - y'), & y = 0 \end{cases} \quad (6)$$

对于概率值较大的简单样本， $(1 - y')$ 经过 γ 参数次方会变得更加小，得到的损失函数值更小，而概率值较小的正样本损失相对较大。负样本同理。因此，Focal loss 能够在网络训练过程中提升难以区分样本的权重，降低简单样本所占的权重。同时，添加平衡因子 α ，用来平衡正负样本比例的不均衡，本文设置平

衡因子 $\alpha = 0.25$ ， $\gamma = 2$ 。

3 实验与分析

光伏电池 EL 图像的裂纹缺陷检测算法是在 Pycharm 上基于 TensorFlow 的深度学习框架进行编程，实验环境是在 Ubuntu 18.04 平台系统上完成，用于训练的计算机是具有 64 GB 内存的 Inter Xeon W 2123 和具有 12 GB 显存的 Titan XP 图形卡。

3.1 数据集标注

数据集中的所有图像都采用 LabelImg 标注软件进行标注。通过使用不同大小的矩形框进行手动标记缺陷位置的同时给定类别标签，制作了 VOC2007 格式的光伏电站 EL 图像数据集，随机分配构成模型的训练集和测试集。表 1 显示了数据集配置。

表 1 光伏电池 EL 图像数据集
Table 1 Photovoltaic cell EL image data set

| 分辨率 | 训练集 | 测试集 | 合计 |
|-----------|-----|-----|-----|
| 1024×1024 | 476 | 236 | 712 |

3.2 实验结果与分析

本文采取平均精度(average precision, AP)来评估检测模型的性能。AP 值的计算与准确率(accuracy, 简称为 P_{pre})、召回率(recall, 简称为 R_{recall})有关，表示的是以召回率作为横坐标，准确率作为纵坐标，绘制的精确/召回率(PR)曲线下的面积^[17]。对于给定的类别，准确率和召回率分别定义为

$$P_{pre} = T_p / (T_p + F_p) \quad (7)$$

$$R_{recall} = T_p / (T_p + F_n) \quad (8)$$

$$AP = \int_0^1 P_{pre}(R_{recall}) dR_{recall} \quad (9)$$

其中： T_p 为分类正确的缺陷目标的个数， F_p 为背景分类为缺陷目标的个数， F_n 是缺陷正样本被错误地分为负样本的个数。

为了更好地评估网络模型的性能，首先对模型参数进行了选取，设置训练迭代次数为 20000，学习率为 0.0001。在 RPN 网络中，生成的 proposal 足够多会在一定程度避免出现缺陷的漏检，但是全部用于后续训练会降低网络的训练速度，加重训练计算负担。因此，需要采用非极大值抑制算法 NMS 完成 proposal 的选择，这里设置 RPN 网络训练时的非极大值抑制阈值参数为 0.7，NMS 之后的 proposal 数量设置为 2000。其他详细参数见表 2。

表 2 模型的参数配置

Table 2 Parameter configuration of the model

| Image_resize | Weight_decay | Learning_rate | Network_batch_size |
|---------------|--------------------------|--------------------------|--------------------|
| 1024×1024 | 0.0005 | 0.0001 | 1 |
| Momentum | RPN_proposals_train | RPN_proposals_test | RPN_batch_size |
| 0.9 | 2000 | 1000 | 256 |
| Max_iteration | ROI_foreground threshold | ROI_background threshold | RPN_nms threshold |
| 20000 | (0.7, 1) | (0, 0.3) | 0.7 |

为进一步观察 ResNet50 与 PA-FPN 结合之后提取的特征，图 6 展示了特征可视化的结果。网络输入为 1024 pixels×1024 pixels 的 EL 图像，经过 ResNet50 残差模块输出特征图 $\{C_2, C_3, C_4, C_5\}$ ，裂纹缺陷区域与背景对比不明显。 $\{P_2, P_3, P_4, P_5\}$ 是 ResNet50 结合 FPN 之后获取的特征图， $\{N_2, N_3, N_4, N_5\}$ 表示通过自下而上

及横向连接再次进行路径聚合后获得的特征图。可以看出，ResNet50 结合 PA-FPN 相较于结合 FPN 所得的特征图中，裂纹区域与复杂背景的分度更加明显，有利于提高网络对多尺度裂纹的特征表达能力。

当 RPN 的输入为图 6 中的特征图时，图 7 展示了 RPN 对特征图进行 3×3 卷积得到的 F 以及融合注意力

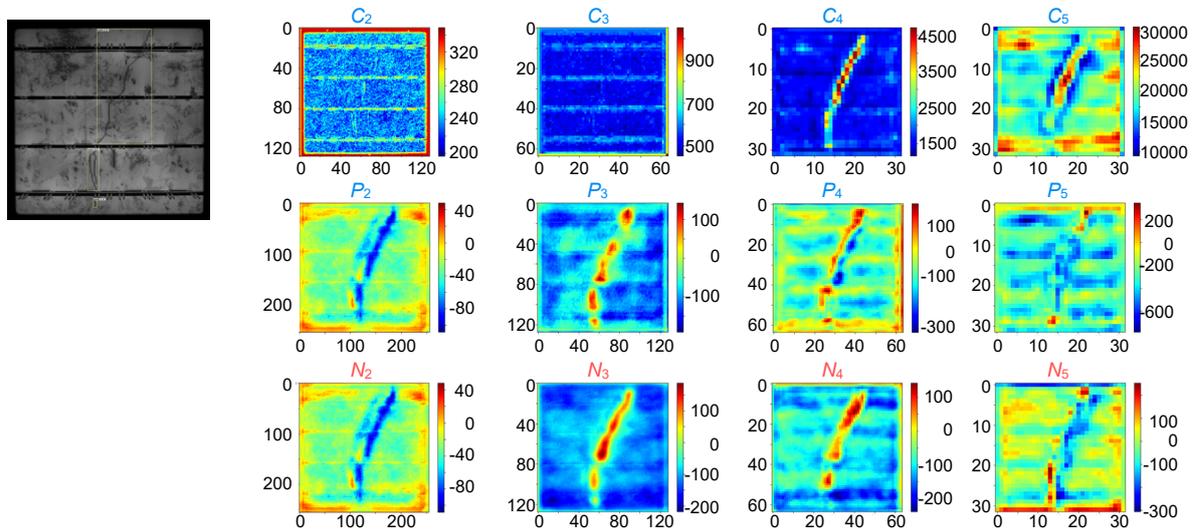


图 6 特征图可视化对比

Fig. 6 Visual comparison of feature maps

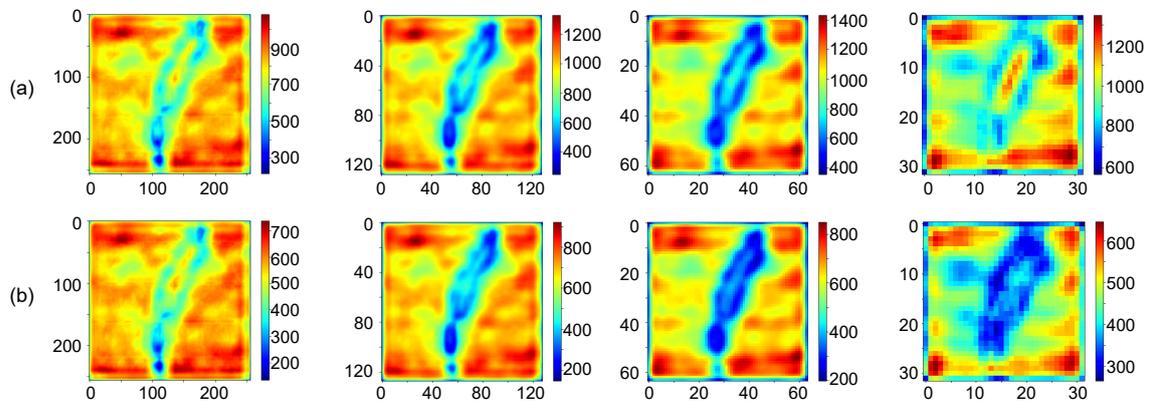


图 7 RPN 结合注意力 CBAM 前后的特征图

Fig. 7 Feature map before and after RPN combined with attention CBAM

CBAM 输出的 F'' 对比图。图 7(a)代表特征图 F ,图 7(b)代表经过 CBAM 输出的 F'' 。图 7 中可以清晰地看出,融合注意力后提取的特征图 F'' 更加鲁棒,模型对裂纹缺陷的关注更高,很好地抑制了复杂背景及晶粒伪缺陷的特征,可以表达更加复杂的语义信息。

为了证明所提方法的有效性,在 EL 数据集上进行了如表 3 所示的实验。由表可知,采用原始的 Faster RCNN 对 EL 图像的多尺度裂纹检测时,最终的 AP 值为 87.68%。在此基础上,RPN 网络结合损失函数 Focal loss, AP 值提升了 1.25%。然后模型融合注意力 CBAM 模块,同时特征提取采用 ResNet 与改进的路径聚合 PA-FPN,模型的 AP 值分别提高了 3.33%和 2.49%,达到 94.75%。这说明,利用注意力 CBAM 关注裂纹特征并抑制复杂背景及其中的晶粒伪缺陷的特征,以及利用 PA-FPN 提取多尺度特征,可以更好地提升对多尺度裂纹尤其是小目标裂纹的检测结果。值得注意的是,在 RPN 与 Focal loss 结合的基础上,将通道注意力与空间注意力融合到 RPN 中的 AP 值分别为 88.94%和 87.92%。这说明,只融合通道注意力对模型的检测效果影响不大,且只融合空间注意力会使模型的检测结果下降 1.01%。再一次验证了文献[18]中提到的:注意力模块 CBAM 优于仅使用其中一种注意力对模型检测结果的影响。

3.3 实验对比

将本文改进的模型与 Faster RCNN、CenterNet、RetinaNet 等模型在光伏电池 EL 图像下的检测效果进

行对比,通过平均准确率 AP 评估其性能,如表 4 所示。

根据实验结果可以看出,改进的 Faster RCNN 检测网络的 AP 值高于原始的二阶段检测网络 Faster RCNN、一阶段检测网络 CenterNet 和 RetinaNet。对于 CenterNet,基于 DLA 特征提取的模型优于基于残差网络 ResNet18 的模型,整体性能与原始的 Faster RCNN 模型性能相近,但由于 CenterNet 模型对于检测小目标以及尺度变化大的目标适应性不是很好,使得模型整体的性能远低于改进的多尺度 Faster RCNN。另外,经典的一阶段检测网络 RetinaNet 比多尺度 Faster RCNN 模型的精度低 10.22%,究其原因,由于 EL 图像具有非均匀复杂纹理特征,背景中包含栅线以及大量的不规则晶粒,而 RetinaNet 无法抑制复杂背景对裂纹检测的干扰,因此对于条状目标的泛化能力差。

图 8 显示了不同检测模型对 EL 图像的对比检测结果。可以看出,对于同一张非均匀纹理复杂背景的光伏电池 EL 图像,CenterNet 的检测结果与原始的 Faster RCNN 接近,但仍然存在小目标的漏检,而且模型对目标框的回归也不够精确。而 RetinaNet 由于无法更好地抑制复杂背景的干扰,且模型对较窄的条状目标适应性不是很好,使其对裂纹缺陷的漏检率较高,模型的检测性能较差。对于融合注意力的多尺度 Faster RCNN 模型,尽管 EL 图像具有非均匀纹理复杂背景,明暗度也不同,裂纹缺陷既包含树状、条状且

表 3 基于 Faster-RCNN 算法的 EL 图像检测性能

Table 3 EL image detection performance based on Faster-RCNN algorithm

| Faster-RCNN | Focal loss | 注意力 | PA-FPN | AP |
|-------------|------------|-----|--------|-------|
| | - | - | - | 87.68 |
| ResNet50 | ✓ | - | - | 88.93 |
| | ✓ | ✓ | - | 92.26 |
| | ✓ | ✓ | ✓ | 94.75 |

表 4 不同算法在光伏电池 EL 图像上的检测性能

Table 4 Detection performance of different algorithms on photovoltaic cell EL images

| Method | 骨干网络 | AP |
|--------------------------|----------|-------|
| 原始 Faster RCNN | ResNet50 | 87.68 |
| | ResNet18 | 85.07 |
| CenterNet ^[5] | DLA | 87.25 |
| | ResNet50 | 84.53 |
| 改进的 Faster RCNN | ResNet50 | 94.75 |

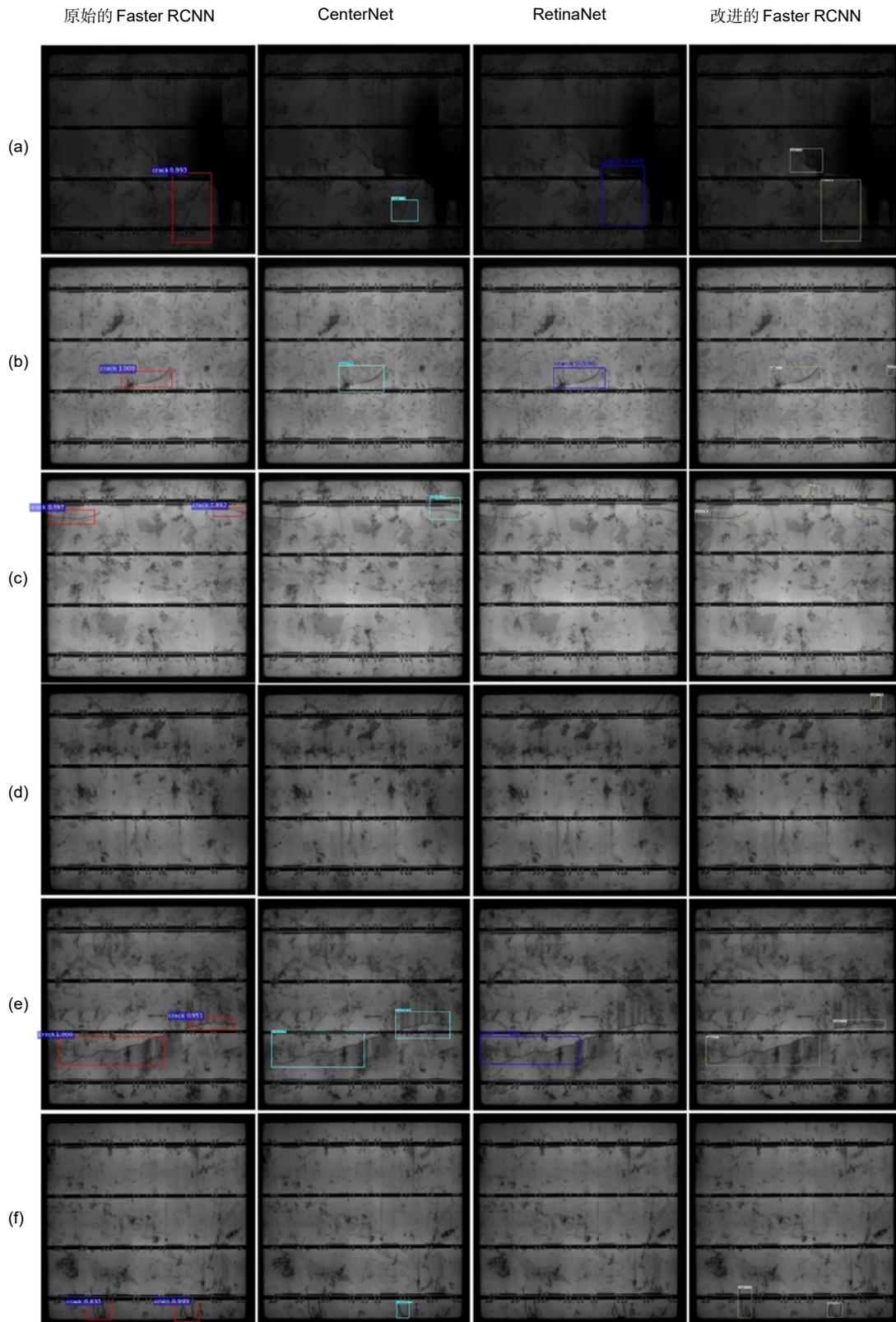


图 8 不同算法在光伏电池 EL 图像上的检测结果对比图

Fig. 8 Comparison of detection results of different algorithms on photovoltaic cell EL images

它们的尺度都不同,改进的模型依旧可以准确地识别出裂纹缺陷,且对它们回归的目标框也较为精准。这是因为,融合注意力的多尺度 Faster R-CNN 在深度残差网络 ResNet50 的基础上,采用了 PA-FPN 结构对多层特征进行融合,获取多尺度的特征图来表达更为复杂的语义信息,对于多尺度裂纹尤其是尺度较小的裂纹检测效果较好。另一方面,RPN 融合了注意力模块 CBAM 与损失函数 Focal loss,提高了网络对裂纹缺陷以及难以区分的样本的比重,提高了模型对缺陷与背景的区别度,很好地提升了检测精度。

总的来说,融合注意力的多尺度 Faster RCNN 一方面抑制了 EL 图像复杂背景对裂纹检测的干扰,另一方面模型对于 EL 图像中尺度变化较大、形状多样且较窄的裂纹缺陷的适应性很好,降低了小目标裂纹漏检的几率,提升了整个模型的检测性能。

4 结 论

基于光伏电池 EL 图像的裂纹缺陷检测,本文提出了融合注意力的多尺度 Faster-RCNN 模型,将 ResNet50 与路径聚合特征金字塔 PA-FPN 相结合,同时采用改进的注意力区域推荐网络 A-RPN 与损失函数 Focal loss,提高了网络在非均匀纹理复杂背景下对多尺度裂纹缺陷的检测能力。相比于几种经典的目标检测算法,改进的模型具有更优的检测性能。本文的研究为光伏电池 EL 图像的多尺度裂纹检测提供了一种较为有效的方法。下一步将继续研究更高精度的目标检测算法,进一步探讨如何增强缺陷特征抑制背景的干扰以提升网络的性能。

参考文献

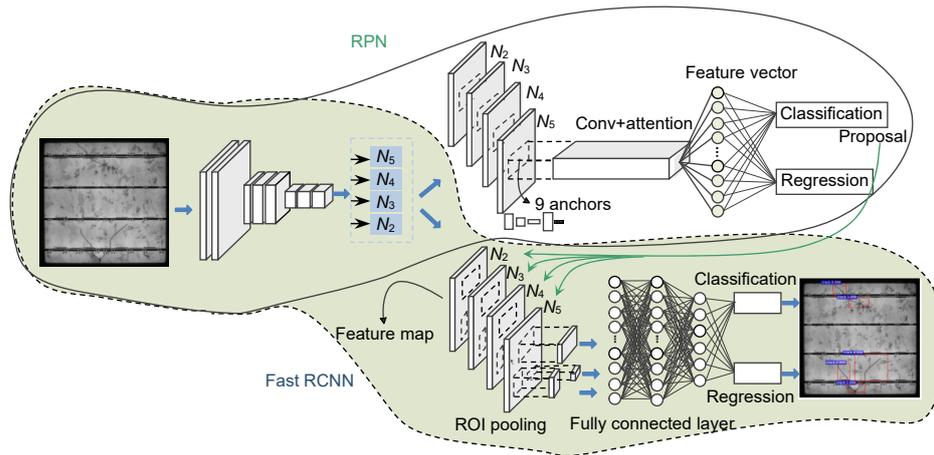
- [1] Anwar S A, Abdullah M Z. Micro-crack detection of multicrystalline solar cells featuring shape analysis and support vector machines[C]//*Proceedings of 2012 IEEE International Conference on Control System, Computing and Engineering*, 2012: 143–148.
- [2] Su B Y, Chen H Y, Zhu Y F, et al. Classification of manufacturing defects in multicrystalline solar cells with novel feature descriptor[J]. *IEEE Trans Instrum Meas*, 2019, **68**(12): 4675–4688.
- [3] Luo Q W, Sun Y C, Li P C, et al. Generalized completed local binary patterns for time-efficient steel surface defect classification[J]. *IEEE Trans Instrum Meas*, 2019, **68**(3): 667–679.
- [4] Tsai D M, Chang C C, Chao S M. Micro-crack inspection in heterogeneously textured solar wafers using anisotropic diffusion[J]. *Image Vis Comput*, 2010, **28**(3): 491–501.
- [5] Cha Y J, Choi W, Büyüköztürk O. Deep learning - based crack damage detection using convolutional neural networks[J]. *Comput Aided Civ Inf Eng*, 2017, **32**(5): 361–378.
- [6] Lin H, Li B, Wang X G, et al. Automated defect inspection of LED chip using deep convolutional neural network[J]. *J Intell Manuf*, 2019, **30**(6): 2525–2534.
- [7] Duan K W, Bai S, Xie L X, et al. Centernet: keypoint triplets for object detection[C]//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*, 2019: 6568–6577.
- [8] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, 2017: 2999–3007.
- [9] Girshick R. Fast R-CNN[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, 2015: 1440–1448.
- [10] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*, 2015: 91–99.
- [11] Cha Y J, Choi W, Suh G, et al. Autonomous structural visual inspection using region - based deep learning for detecting multiple damage types[J]. *Comput Aided Civ Inf Eng*, 2018, **33**(9): 731–747.
- [12] Gao L, Chen N N, Fan Y. Vehicle detection based on fusing multi-scale context convolution features[J]. *Opto-Electron Eng*, 2019, **46**(4): 180331.
高琳, 陈念年, 范勇. 融合多尺度上下文卷积特征的车辆目标检测[J]. *光电工程*, 2019, **46**(4): 180331.
- [13] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 8759–8768.
- [14] Corbetta M, Shulman G L. Control of goal-directed and stimulus-driven attention in the brain[J]. *Nat Rev Neurosci*, 2002, **3**(3): 201–215.
- [15] Frazão M, Silva J A, Lobato K, et al. Electroluminescence of silicon solar cells using a consumer grade digital camera[J]. *Measurement*, 2017, **99**: 7–12.
- [16] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 7132–7141.
- [17] Everingham M, Van Gool L, Williams C K I, et al. The PASCAL visual object classes (VOC) challenge[J]. *Int J Comput Vis*, 2010, **88**(2): 303–338.
- [18] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the European Conference on Computer Vision (ECCV)*, 2018: 3–19.

Crack detection based on multi-scale Faster RCNN with attention

Chen Haiyong^{1*}, Zhao Peng¹, Yan Haowei²

¹School of Artificial Intelligence, Hebei University of Technology, Tianjin 300000, China;

²Tianjin Aerospace Zhongwei Data System Technology Co., Ltd, Tianjin 300000, China



Multiscale crack detection network structure

Overview: Electroluminescence (EL) images of photovoltaic cells have a non-uniformly textured complex background, and the background contains grain pseudo-defects that are highly similar to the crack structure. At the same time, the cracks are characterized by various sizes and shapes. Existing target detection algorithms based on convolutional neural networks cannot adapt to the above problems. From the perspective of suppressing interference from complex background and improving the adaptability of the model to multi-scale crack defect detection, this paper proposes a multi-scale Faster RCNN model that integrates attention. In photovoltaic cell EL images, the scale of the cracks varies greatly, including a large number of small target cracks. In order to improve the network's ability to express multi-scale crack defects, a path aggregation feature pyramid network (PA-FPN) is proposed. Based on the combination of the residual network ResNet50 and the feature pyramid network FPN, PA-FPN adds a bottom-up path to fuse features. PA-FPN effectively retains shallow feature information, which improves the model's adaptability to multi-scale cracks in EL images and especially the detection results of small-scale cracks. In order to improve the model's attention to crack defects and suppress the characteristics of complex background and grain pseudo-defects, this paper proposes a regional recommendation network A-RPN that incorporates convolutional block attention module (CBAM). CBAM is composed of a channel attention module and a spatial attention module. In this paper, it is experimentally verified that the detection result of the RPN network fused with CBAM is better than that of using an attention modules alone. K-means clustering is used to cluster the crack sizes in the data set to guide the RPN to set the anchor box closer to the actual crack size, which improves the speed and accuracy of the target box regression in the defect detection process. In addition, in the RPN network training process, the loss function Focal loss is used to replace the original cross-entropy loss function, so as to reduce the proportion of simple samples in the training process and make the model pay more attention to the samples that are difficult to distinguish. The entire network can achieve end-to-end training. In order to verify the effectiveness of the improved algorithm, the performance of the original Faster RCNN model, RetinaNet, and CenterNet on multi-scale crack detection of EL images is compared. Through training and testing of 1024 pixels×1024 pixels of photovoltaic cell EL images, experimental results show that the improved Faster RCNN is better than the above mentioned target detection algorithms in accuracy, and has good robustness to the strip-shaped multi-scale cracks, which can be adapted to the EL image with changing complex background.

Chen H Y, Zhao P, Yan H W. Crack detection based on multi-scale Faster RCNN with attention[J]. *Opto-Electron Eng*, 2021, 48(1): 200112; DOI: 10.12086/oe.2021.200112

Foundation item: National Natural Science Foundation of China (61873315)

* E-mail: haiyong.chen@hebut.edu.cn