

# 光电工程

## Opto-Electronic Engineering

中文核心期刊 中国科技核心期刊  
Scopus CSCD

### 局部和全局特征融合的太阳能电池片表面缺陷检测

陶志勇, 何燕, 林森, 易廷军, 张尧晟

#### 引用本文:

陶志勇, 何燕, 林森, 等. 局部和全局特征融合的太阳能电池片表面缺陷检测[J]. *光电工程*, 2024, 51(1): 230292.

Tao Z Y, He Y, Lin S, et al. Surface defect detection of solar cells using local and global feature fusion[J]. *Opto-Electron Eng*, 2024, 51(1): 230292.

<https://doi.org/10.12086/oe.2024.230292>

收稿日期: 2023-12-01; 修改日期: 2024-02-02; 录用日期: 2024-02-02

### 相关论文

#### 面向道路场景语义分割的移动窗口变换神经网络设计

杭昊, 黄影平, 张栩瑞, 罗鑫

*光电工程* 2024, 51(1): 230304 doi: 10.12086/oe.2024.230304

#### 融合多分辨率特征的点云分类与分割网络

陶志勇, 李衡, 豆淼森, 林森

*光电工程* 2023, 50(10): 230166 doi: 10.12086/oe.2023.230166

#### 融合Swin Transformer的立体匹配方法STransMNet

王高平, 李珣, 贾雪芳, 李哲文, 王文杰

*光电工程* 2023, 50(4): 220246 doi: 10.12086/oe.2023.220246

#### 融合注意力的多尺度Faster RCNN的裂纹检测

陈海永, 赵鹏, 闫皓炜

*光电工程* 2021, 48(1): 200112 doi: 10.12086/oe.2021.200112

更多相关论文见光电期刊集群网站 



<http://cn.ojournal.org/oe>



 OE\_Journal



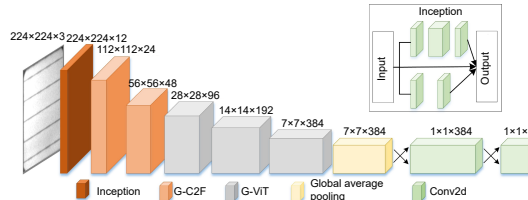
Website

DOI: 10.12086/oe.2024.230292

# 局部和全局特征融合的太阳能电池片表面缺陷检测

陶志勇<sup>1</sup>, 何燕<sup>1\*</sup>, 林森<sup>2</sup>, 易廷军<sup>1</sup>, 张尧晟<sup>1</sup><sup>1</sup> 辽宁工程技术大学电子与信息工程学院, 辽宁 葫芦岛 125105;<sup>2</sup> 沈阳理工大学自动化与电气工程学院, 辽宁 沈阳 110159

**摘要:** 太阳能电池片表面缺陷具有类内差异大、类间差异小和背景特征复杂等特点, 因此, 要实现高精度的太阳能电池片表面缺陷自动检测是一项富有挑战性的任务。针对此问题, 该文提出融合局部和全局特征的卷积视觉Transformer网络(CViT-Net), 首先采用Ghost聚焦(G-C2F)模块提取电池片缺陷局部特征; 然后引进坐标注意力强调缺陷特征并抑制背景特征; 最后构建Ghost视觉(G-ViT)模块融合电池片缺陷局部特征和全局特征。同时, 针对不同检测精度和模型参数量, 分别提供了CViT-Net-S和CViT-Net-L两种网络结构。实验结果表明, 与经典MobileVit、MobileNetV3和GhostNet轻量级网络相比, CViT-Net-S对电池片分类准确率分别提升了1.4%、2.3%和1.3%, 对电池片检测mAP50分别提升了2.7%、0.3%和0.8%; 与ResNet50、RegNet网络相比, CViT-Net-L分类准确率分别提升了0.72%和0.7%, 检测mAP50分别提升了3.9%、1.3%; 与先进YOLOv6、YOLOv7和YOLOv8检测网络相比, 作为骨干网络的CViT-Net-S、CViT-Net-L结构在mAP和mAP50指标上仍保持良好检测效果。结果证明本文算法在太阳能电池片表面缺陷检测领域具有应用价值。

**关键词:** 深度学习; 特征融合; 太阳能电池; 缺陷分类; 缺陷检测**中图分类号:** TP391.4; TP274**文献标志码:** A

陶志勇, 何燕, 林森, 等. 局部和全局特征融合的太阳能电池片表面缺陷检测[J]. 光电工程, 2024, 51(1): 230292

Tao Z Y, He Y, Lin S, et al. Surface defect detection of solar cells using local and global feature fusion[J]. *Opto-Electron Eng*, 2024, 51(1): 230292

## Surface defect detection of solar cells using local and global feature fusion

Tao Zhiyong<sup>1</sup>, He Yan<sup>1\*</sup>, Lin Sen<sup>2</sup>, Yi Tingjun<sup>1</sup>, Zhang Yaosheng<sup>1</sup><sup>1</sup> School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China;<sup>2</sup> School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, Liaoning 110159, China

**Abstract:** The surface defects of solar cells exhibit significant intra-class differences, minor inter-class differences, and complex background features, making high-precision identification of surface defects a challenging task. This paper proposes a Convolutional-Vision Transformer Network (CViT-Net) that combines local and global features to address this issue. First, a Ghost-Convolution two-fusion (G-C2F) module is used to extract local features of the solar cell panel defects. Then, a coordinate attention mechanism is introduced to emphasize defect features and

收稿日期: 2023-12-01; 修回日期: 2024-02-02; 录用日期: 2024-02-02

基金项目: 辽宁省科技厅应用基础研究项目(2022JH2/101300274); 辽宁省研究生教育教学改革研究项目(LNYJG2023117); 辽宁省教育厅基本科研项目(LJKMZ20220676)

\*通信作者: 何燕, 2575561115@qq.com。

版权所有©2024 中国科学院光电技术研究所

suppress background features. Finally, a Ghost-Vision Transformer (G-ViT) module is constructed to fuse local and global features of the solar cell panel defects. Meanwhile, CViT-Net-S and CViT-Net-L network structures are provided for low-resource and high-resource environments. Experimental results show that compared to classic lightweight networks such as MobileVit, MobileNetV3, and GhostNet, CViT-Net-S improves the classification accuracy of solar cell panels by 1.4%, 2.3%, and 1.3%, respectively, and improves the mAP50 for defect detection by 2.7%, 0.3%, and 0.8% respectively. Compared to ResNet50 and RegNet, CViT-Net-L enhances the classification accuracy by 0.72% and 0.7%, respectively, and improves the mAP50 for defect detection by 3.9% and 1.3%, respectively. Compared to advanced YOLOV6, YOLOV7, and YOLOV8 detection networks, CViT-Net-S and CViT-Net-L structures, as backbone networks, still maintain good detection performance in terms of mAP and mAP50 metrics, demonstrating the application value of the proposed algorithm in the field of solar cell panel surface defect detection.

**Keywords:** deep learning; feature fusion; solar cells; defect classification; defect detection

## 1 引言

太阳能<sup>[1-2]</sup>是绿色环保清洁能源,具有无污染和可再生特点。单晶硅太阳能电池片具有稳定性好、材料分布广、能量转换效率高的优势,在工业制造中电池片通过串联、并联焊接的太阳能电池组件作为太阳辐射能转换为电能的核心部件,能量转换效率的高低和太阳能电池片表面缺陷程度密切相关。为提高电池组件的良品率,避免隐裂、暗斑、瑕疵和黑心等缺陷电池片进入电池组件,需要对电池片进行表面缺陷类型和位置的长期监控,以剔除缺陷电池片并提高产品的合格率。

电池片表面缺陷的识别有传统的机器学习方法和深度学习方法。传统的机器学习方法<sup>[3]</sup>通过人工进行电池片表面缺陷特征提取,然后使用传统算法对图像纹理特征进行分类。经典方法有 Juan 等人<sup>[4]</sup>提出的电致发光成像联合支持向量机 (support vector machines, SVM) 分类器的检测技术,它通过对太阳能电池片缺陷特征、发光特性和健康状况等缺陷识别,多角度分析缺陷电池片对光伏系统输出效率的影响;王超等人<sup>[5]</sup>通过分析光伏电池缺陷和扩散长度的关系,解决了隐裂缺陷的可视化识别;Firuzi 等人<sup>[6]</sup>利用定向梯度直方图 (histogram of oriented gradient, HOG) 结合局部二进制模式 (local binary patterns, LBP) 进行缺陷识别,取得了较好的效果;Kim 等人<sup>[7]</sup>利用 K 最近邻 (K-nearest neighbor, KNN) 算法和图像处理技术进行光伏电池片缺陷自动识别,解决电池片缺陷识别速度慢的问题。尽管传统机器学习方法在电池片表面缺陷检测中具有识别效果好、算法成熟的优点,但仍有算法需要大量参数调整、模型鲁棒性差、模型泛化

性能差、依赖工程师主观性经验判别和无法长时间人工作业等问题。

深度学习<sup>[8]</sup>的电池片缺陷识别技术包含缺陷分类和缺陷检测。针对电池片中单类别缺陷,部分研究者使用红外图片、电致发光图片的缺陷分类技术。Masita 等人<sup>[9]</sup>结合 ResNet50 残差网络<sup>[10]</sup>来处理电池片热差,通过红外图像热差来判别电池片缺陷;Du 等人<sup>[11]</sup>结合传统分类检测和深度学习智能分类检测进行电致发光隐裂缺陷识别;Chen 等人<sup>[12]</sup>利用卷积神经网络提取电池片图像光谱特征来分析缺陷图像的复杂纹理信息,实现了较好的缺陷图像分类准确率。尽管深度学习方法在电池片缺陷分类技术中取得了良好的识别效果,但对于电池片中的多类别缺陷,研究者们仍在不断努力。Su 等人<sup>[13]</sup>通过改进 Faster R-CNN 网络<sup>[14]</sup>的区域建议框结构,解决多晶硅电池片中栅、隐裂缺陷平均检测准确率不高的问题;Zhang 等人<sup>[15]</sup>通过添加 YOLOv5 网络<sup>[16]</sup>的多尺度检测头和聚类优化先验框的思路,解决电池片定位精度不高和检测速度不快的问题。尽管这些网络对电池片特定缺陷识别效果良好,由于缺陷特征相似度高、背景特征复杂,致使电池片细粒度缺陷特征提取不充分,以及网络加深时易出现特征丢失现象,从而导致检测精度低。

为解决网络对特征提取不充分和检测精度低的问题,结合深度学习先进技术,提出高质量融合局部和全局特征的卷积视觉自注意力网络 (convolutional-vision Transformer networks, CViT-Net)。该网络拥有缺陷特征提取能力强、缺陷分类检测精度高等优势。本文主要贡献如下:

1) 提出了两个新颖模块分别为 Ghost 聚焦 (Ghost-

convolution two fusion, G-C2F) 模块和鬼影视觉 (Ghost-vision Transformer, G-ViT) 模块, G-C2F 旨在提取电池片缺陷局部特征, 解决局部细粒度信息提取不充分的问题; G-ViT 旨在融合电池片局部特征和全局特征, 解决模型对细粒度信息的特征丢失问题。

2) 提出了融合局部特征和全局特征的 CViT-Net 网络, 该网络具有特征表达能力强的优势; 为适应不同工业场景, 设计了 CViT-Net-S 和 CViT-Net-L, 分别针对低和高检测精度, 以满足太阳能电池缺陷检测的不同需求。

3) CViT-Net 在电池片缺陷分类和缺陷检测任务中具有良好识别效果, CViT-Net-S 结构和 CViT-Net-L 结构相较经典轻量级分类模型具有特征提取能力强、分类准确率高的特性; 相较先进 YOLO 系列模型仍保持较好的缺陷检测能力和缺陷定位能力。

## 2 相关工作

### 2.1 轻量级图像分类模型

近年来, 提高神经网络架构的检测精度成为计算机视觉的重要挑战, 轻量化模型以较少的参数量成为视觉领域的研究热点。标准的卷积神经网络包含卷积层、池化层和非线性激活函数, 由于计算成本大和空间复杂度高, 很多网络通过改善标准卷积的计算方法来降低模型复杂度和计算成本。其中 MobileNetV3<sup>[17]</sup> 在反向倒残差结构中融合深度可分离卷积实现模型轻量化, ShuffleNetV2<sup>[18]</sup> 结合组卷积和通道洗牌策略降低内存和模型参数量, EfficientNet<sup>[19]</sup> 综合分析模型深度、宽度和分辨率实现少参数量的高精度网络, GhostNet<sup>[20]</sup> 对输入特征图简单线性变换构成鬼影 (Ghost) 特征图后进行叠加降低模型复杂度, RegNet<sup>[21]</sup> 融合组卷积和残差架构分支设计不同计算量的网络。这些网络可以高效地替换特定任务中骨干网络 (如 ResNet50), 以减少模型规模并改善延迟。但模型存在特征空间信息提取局限的问题, 不利于捕获图像的全局特征信息。

为高效捕获图像全局特征信息, 将视觉 Transformer (vision Transformer, ViT)<sup>[22]</sup> 的特征提取网络应用在图像识别领域, 对大规模数据集和高性能训练模式, ViT 在没有图像特定诱导偏差下可实现卷积神经网络同级别的识别精度。然而, ViT 本身具有缺乏归纳偏差的固有属性, 使得模型难以在轻量和高精度之间取得平衡, 为了提升模型的识别性能, 在 ViT 中引入卷

积用于整合这种归纳偏差。CvT 模型<sup>[23]</sup> 在 ViT 模型的多头自注意力中引用深度可分离卷积; ConViT 模型<sup>[24]</sup> 学习卷积神经网络的局部相关性来摆脱卷积的局部特征; MobileViT<sup>[25]</sup> 将提取的局部特征进行全局表达以获取丰富的图像细粒度特征。

### 2.2 经典目标检测模型

在计算机视觉识别任务中, 目标检测是一个重要的分支。模型根据特征提取方式, 可分为一阶段检测和二阶段检测。基于回归预测的一阶段检测模型 (如 YOLO) 拥有检测速度快、效率高的优势, 二阶段检测模型 (如 Faster R-CNN) 具有检测精度高、检测速度慢的特点。一阶段检测模型中, 其原理是在模型训练之前进行聚类优化锚框来提高检测精度的有 YOLOv5<sup>[16]</sup>、YOLOv7<sup>[26]</sup>、YOLOR<sup>[27]</sup>; 抛弃锚框采用锚点到目标边界框距离的有 YOLOv6<sup>[28]</sup>、YOLOv8<sup>[29]</sup>、YOLOX<sup>[30]</sup>; Faster R-CNN<sup>[14]</sup> 等二阶段检测模型原理是通过生成区域建议框来提取图像特征, 然后对这些提取的特征进行目标分类和目标预测。

一、二阶段检测模型的网络结构主要由骨干、颈部、头部网络三部分组成。骨干网络用于提取图像特征, 为满足不同检测需求和提高检测精度, 研究人员微调 DenseNet<sup>[31]</sup> 和 MobileNetV3<sup>[17]</sup> 模型后作为检测模型的骨干网络, 以增强特征提取能力; 颈部网络用于提取整理骨干网络提取的丰富特征信息, 采用 FPN<sup>[32]</sup> 和 PANet<sup>[33]</sup> 两种方式来聚合图像的细粒度特征; 头部网络则负责预测目标类别和目标边界框, 主要分为二阶段检测方法 (Faster R-CNN) 和一阶段检测方法 (YOLO) 两种形式。

## 3 卷积视觉 Transformer 网络

### 3.1 坐标注意力

坐标注意力 (Coordinate attention, CA)<sup>[34]</sup> 是一种增强模型特征提取能力的计算单元, 在不增加网络复杂度的情况下突出特征信息。压缩激发 (Squeeze-and-excitation, SE)<sup>[35]</sup> 注意力利用全局池化进行全局空间信息编码, 编码后的空间信息被压缩到通道信息中, 位置信息难以保存。空间注意力 (Convolutional block attention module, CBAM)<sup>[36]</sup> 通过考虑特征图空间和通道维度之间的语义相互依赖性来建立跨通道和跨空间的信息, 但由于引入大尺度卷积增加计算开销。通道和空间兼有的多尺度注意力机制 (efficient multi-scale attention, EMA)<sup>[37]</sup>, 由于其较高的参数量和计

算复杂度, 无法满足行业实时检测的需求。坐标注意力不仅考虑了特征间不同位置的相关性, 也分析特征位置之间的距离和方向信息, 结构如图 1 所示。

坐标注意力对输入  $X$  分别使用尺寸为  $(H,1)$  或  $(1,W)$  的全局平均池化核,  $H$  是特征图高度,  $W$  是特征图宽度。沿着水平坐标和垂直坐标对每个通道进行编码, 高度为  $h$  的第  $C$  通道输出  $Z_C^h(h)$  表示为式 (1),  $X_C(h,i)$  表示输入特征  $X_C$  在  $(h,i)$  坐标上的数值。

$$Z_C^h(h) = \frac{1}{W} \sum_{0 \leq i < W} X_C(h,i). \quad (1)$$

宽度为  $w$  的第  $C$  通道输出表示为式 (2),  $X_C(j,w)$  表示输入特征  $X_C$  在  $(j,w)$  坐标上的数值。

$$Z_C^w(w) = \frac{1}{H} \sum_{0 \leq j < H} X_C(j,w). \quad (2)$$

坐标信息嵌入阶段获取的方向感知特征图  $Z_C^h(h)$  和  $Z_C^w(w)$ , 如式 (3) 所示, 方向感知特征图在通道上首先进行拼接操作聚合感兴趣的区域特征, 并利用  $1 \times 1$  卷积函数  $F_1$  进行变换操作得到水平方向和垂直方向的特征映射  $f \in R^{C/r \times (H+W)}$ ,  $r$  是缩减率,  $\delta$  为 ReLU 激活函数。

$$f = \delta(F_1(\text{Concat}[Z_C^h, Z_C^w])). \quad (3)$$

$f$  沿着空间维度分别分解为两个独立的张量  $f^h \in R^{C/r \times H}$  和  $f^w \in R^{C/r \times W}$ ,  $f^h$  和  $f^w$  分别经过卷积核大小为  $1 \times 1$  的卷积函数  $F_h$  和  $F_w$ , 卷积函数生成与  $X$  同样尺寸的特征张量  $g^h$  和  $g^w$ , 如式 (4) 和 (5) 所示,  $\sigma$  表示 Sigmoid 激活函数。

$$g^h = \sigma(F_h(f^h)), \quad (4)$$

$$g^w = \sigma(F_w(f^w)). \quad (5)$$

$g^h$  和  $g^w$  作为坐标注意力权重与原始输入特征  $X$  进行点乘, 输出第  $C$  通道上坐标  $(i,j)$  的增强特征  $Y$  可写

为式 (6)。

$$Y_C(i,j) = X_C(i,j) \times g_C^h(i) \times g_C^w(j). \quad (6)$$

### 3.2 Ghost 聚焦模块

卷积神经网络 (Convolutional neural networks, CNN) 参数共享和并行计算对计算机视觉任务提供了友好发展, 但模型参数量和复杂度会随着网络层数的加深出现非线性增加, 高参数量和高复杂度模型不利于工业进行实时检测, 因此, 同时兼备轻量级和高时效的网络对工业检测具有重要应用价值。该文网络受到轻量级 MobileNetV3<sup>[17]</sup> 和 GhostNet<sup>[20]</sup> 模型启发, 构建了能提取丰富特征表达能力的 Ghost 聚焦 (Ghost-convolution two fusion, G-C2F) 模块, 结构如图 2 所示, G-C2F 模块对特征  $X$  分别输入 CNN 分支和 C2F 分支提取缺陷信息。

G-C2F 模块对输入特征  $X \in R^{H \times W \times C}$  分别经过 CNN 分支和 C2F 分支得到纹理特征  $X_{GC}$  和  $X_{C2F}$ , 然后将  $X_{GC}$  和  $X_{C2F}$  进行融合叠加得到输出特征  $Y \in R^{H \times W \times C}$ , 过程如式 (7) 所示:

$$Y = X_{GC} + X_{C2F}. \quad (7)$$

CNN 分支旨在实现模型轻量化和高精度, 输入张量  $X$  利用 Ghost  $3 \times 3$  卷积提取电池片缺陷特征, 提取后的特征通过 CA 强化缺陷信息, 最后是一个卷积核大小为  $1 \times 1$  的点卷积层学习输入通道的线性组合并将张量投影至高维空间; 同时, 考虑到太阳能电池片纹理信息复杂、类内差异大和类间差异小的特点, 模块引入 ResNet50 模型<sup>[10-11]</sup> 残差架构思想来避免 CViT-Net 网络梯度消失问题。请注意, 本文引入的残差分支和 ResNet50 模型的残差分支存在差异, 在层与层之间, 采用池化核大小为  $3 \times 3$  的最大池化层替代 ResNet50 残差模块的残差分支, 这样做的优势可降低模型过拟合及获取缺陷高频特征, 最后融合最大池化残差分支后的输出张量为  $X_{GC}$ 。

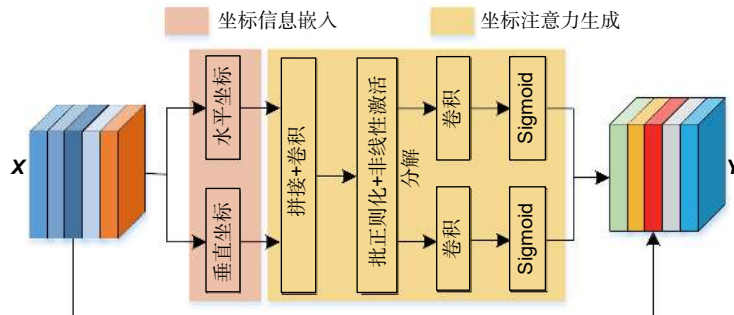


图 1 坐标注意力

Fig. 1 Coordinate attention

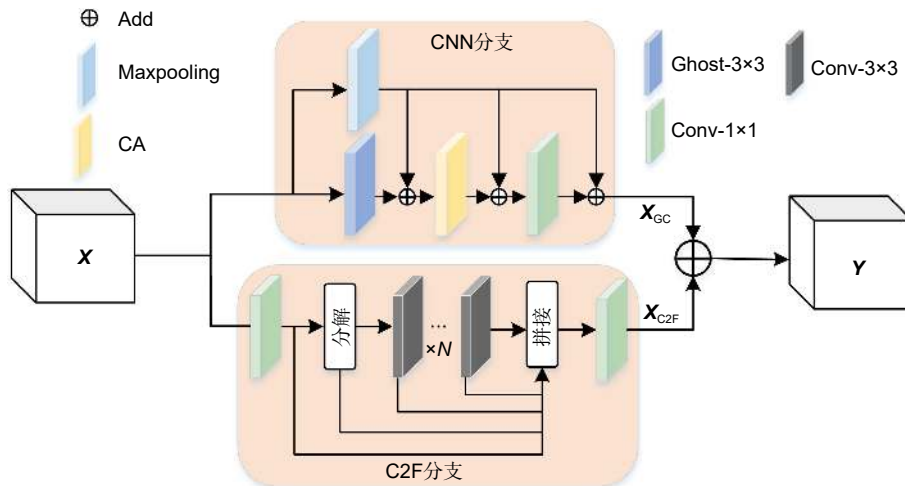


图 2 鬼影聚焦模块  
Fig. 2 Ghost focus module

C2F 分支旨在用卷积层提取丰富的局部细粒度特征，该分支灵感来自于 YOLOv8 网络<sup>[29]</sup>。为实现网络轻量化和丰富的局部图像编码信息，C2F 分支首先对输入张量  $X$  采用卷积核大小为  $1 \times 1$  的点卷积调整缺陷特征编码维度；其次对缺陷特征在空间维度进行细致划分，借鉴了 ELAN<sup>[26]</sup> 架构的方式来捕捉丰富的梯度流信息。然后采用了将模型获取的特征图在通道上进行拼接的方法，以降低参数量和计算量。最后将获取的层级特征进行通道叠加后送入卷积核大小为  $1 \times 1$  的点卷积得到输出特征张量  $X_{C2F}$ 。尽管 C2F 分支中  $N$  有多种取值 (如 2,3,4,5,...)，但从模型参数量和复杂度考虑，发现  $N=3$  时 CViT-Net 模型可实现最佳效果。

### 3.3 鬼影视觉模块

卷积神经网络强归纳偏置对临近像素的局部相关性拥有友好关注度，相比之下，利用自注意力机制捕

获全局语义信息的 Transformer 模型<sup>[22]</sup> 最小化图像内的归纳偏差，但在全局表征学习中它的长距离依赖关系对特征信息提升效果显著。因此，为获取丰富的图像局部纹理特征和全局语义信息，融合 CNN 局部强归纳偏置优势和 Transformer 长距离依赖强项构造了鬼影视觉 (Ghost-vision Transformer, G-ViT) 模块，结构如图 3 所示。

G-ViT 模块旨在对较少输入参数的输入张量进行局部信息和全局信息建模，结构上，对于给定输入张量  $X \in R^{H \times W \times C}$  首先在空间维度上通过分解操作得到两个相同的独立张量  $X_1 \in R^{H \times W \times C/2}$ ， $H$  表示特征图高度， $W$  表示特征图宽度， $C$  是通道数；然后将独立张量  $X_1$  分别经过 CNN 分支编码局部空间信息得到局部特征  $X_{GC}$ ，Transformers 分支进行全局信息建模得到全局特征  $X_{ViT}$ ；最后利用拼接实现局部特征和全局信息聚

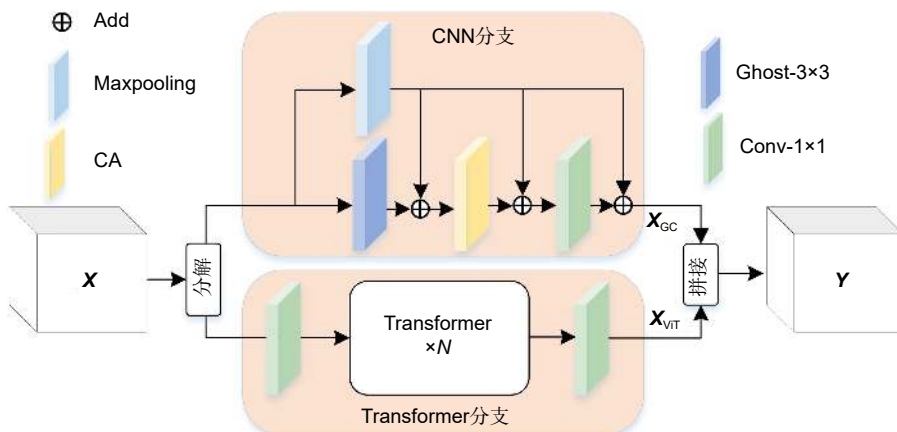


图 3 鬼影视觉模块  
Fig. 3 Ghost vision module

合得到输出特征  $Y \in R^{H \times W \times C}$ , 过程如式 (8) 所示。请注意, G-ViT 模块和 G-C2F 模块的 CNN 分支架构相同, 输出特征依然是用  $X_{GC}$  表示。

$$Y = Concat([X_{GC}, X_{ViT}]) \quad (8)$$

Transformer 分支旨在实现长期建模的非局部依赖关系, 同时获取有效的感知域  $H \times W$ 。首先利用卷积核大小为  $1 \times 1$  的点卷积, 对输入张量  $X_i$  的张量维度进行调整; 其次利用多头自注意力 Transformer 模块捕获丰富的全局缺陷特征, 由于 Transformer 重量级因素不利于模型优化, 因此, 考虑到模型复杂度特点, 模块采用  $N=3$  和头部数量为 8 的 Transformer 结构获取全局信息表示; 最后将获取的全局信息再次通过  $1 \times 1$  的点卷积进行局部空间编码, 得到张量  $X_{ViT}$ 。

### 3.4 CViT-Net 结构

CViT-Net 网络为不同工业实践场景提供了 2 种型号的结构, 分别定义为 CViT-Net-S 和 CViT-Net-L, CViT-Net-S 网络是针对低检测精度 (如移动端), 模型

参数量和计算量分别是 5.6 M 和 1.52 G; CViT-Net-L 网络主要用在高检测精度 (如云端), 模型参数量和计算量分别是 21.9 M 和 6.49 G, CViT-Net 网络结构图如图 4 所示。

详细参数如表 1 所示。这些模型首先通过 Inception<sup>[38]</sup> 模块进行多尺度特征聚合; 然后通过多层次 G-C2F 模块和 G-ViT 模块提取电池片表面缺陷特征, CViT-Net 网络首先通过 G-C2F 模块获取丰富的局部细粒度特征, G-ViT 模块融合局部特征和全局信息; 最后为避免模型过拟合和缺陷局部结构信息丢失, 受到 MobileNetV3 网络<sup>[17]</sup> 启发, 利用卷积核大小为  $1 \times 1$  的卷积层和池化核为  $7 \times 7$  的全局平均池化取代全连接层 (fully connected layers, FC), 并进行缺陷特征分类。分类使用的卷积层不带偏置项和激活函数, 从而最大限度减少参数量和降低计算成本。

CViT-Net-S 网络和 CViT-Net-L 网络输入分辨率都是  $224 \times 224$ , 层级间相同模块输出分辨率的缺陷特

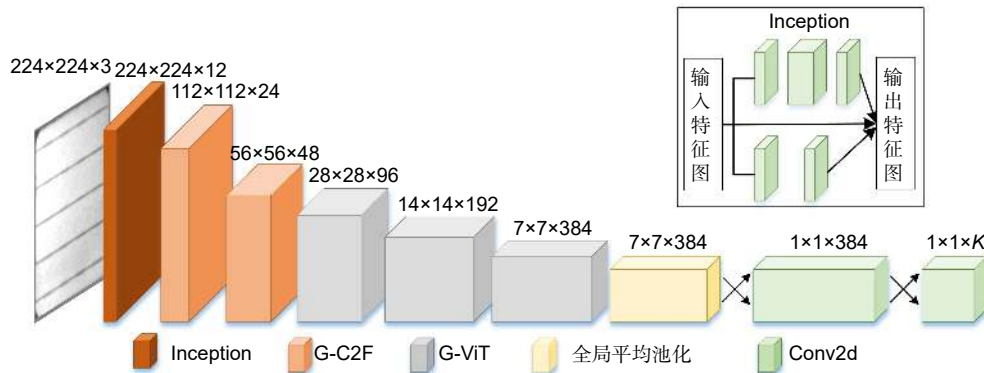


图 4 CViT-Net 网络结构图

Fig. 4 CViT-Net network structure diagram

表 1 CViT-Net 模型参数表

Table 1 CViT-Net model parameter table

输入分辨率	模块	输出分辨率	CViT-Net-S		CViT-Net-L	
			输出通道数	重复	输出通道数	重复
$224 \times 224$	Inception	$224 \times 224$	12	1	24	1
$224 \times 224$	G-C2F	$112 \times 112$	24	1	48	2
$112 \times 112$	G-C2F	$56 \times 56$	48	2	96	2
$56 \times 56$	G-ViT	$28 \times 28$	96	2	192	2
$28 \times 28$	G-ViT	$14 \times 14$	192	4	384	4
$14 \times 14$	G-ViT	$7 \times 7$	384	2	768	2
$7 \times 7$	池化	$1 \times 1$	384	1	768	1
$1 \times 1$	Conv2d	$1 \times 1$	384	1	768	1
$1 \times 1$	Conv2d	$1 \times 1$	K	1	K	1
	Parameter		5.6 M		21.9 M	
	FLOPs		1.52 G		6.49 G	

征大小保持尺寸一致, 它们在池化后经过 2 个卷积核大小是  $1 \times 1$  的点卷积进行缺陷识别分类。然而, 针对不同资源用例的工业场景, CViT-Net-S 网络层级间模块输出通道数和模块重复数与 CViT-Net-L 网络存在差异, 最后经过  $K$  值实现缺陷特征分类。

## 4 数据集和缺陷检测流程

### 4.1 数据集

自建的太阳能电池片数据集采用电致发光 (EL) 成像技术<sup>[39]</sup>, 通过 CCD 相机采集  $6 \times 10$  的电池组件, 进行切割、汇编、标注后构建太阳能电池片图像数据库。本文重点对工业常出现的无缺陷、黑心、暗斑、瑕疵和隐裂 5 类进行研究, 所有检测图像分辨率为  $540 \times 540$ , 类型如图 5 所示。无缺陷表现出电池片正常完好无损; 黑心表现出电池片是整体黑色的图片; 暗斑表现为图像中呈块状型深暗色或黑色的电池片; 瑕疵表现为电池片中微小的斑点状或小黑点; 隐裂表现为电池边缘与栅线附近的延伸裂开状缺陷。

### 4.2 缺陷检测流程

电池片图像包含多种类型缺陷、缺陷背景纹理特征复杂、类内差异大和类间差异小的问题。因此将数

据库中的电池片种类图像划分为分类数据集和检测数据集。分类数据集有三个类别, 分别是无缺陷、黑心和多类缺陷图像 (隐裂、暗斑、瑕疵), 数据集中图像均为原始图像, 未做任何清晰化等图像预处理, 总计包含数据量是 9790 片, 利用分类模型 (CViT-Net) 识别出无缺陷、黑心和多类别缺陷; 检测数据集是在分类数据集的基础上存在的多类缺陷图像 (隐裂、暗斑、瑕疵) 进行边界框的定位检测, 总计数据量是 4970 片, 利用定位模型 YOLOv5(CViT-Net) 定位出图像内隐裂、暗斑和瑕疵缺陷的位置, YOLOv5(CViT-Net) 表示采用 CViT-Net 为检测模型骨干网络, YOLOv5 为检测框架。详细的缺陷检测流程图如图 6 所示。为保证网络合理的训练和测试, 将电池分类数据集和检测数据集分别按照 6:2:2 比例划分为训练集、验证集、测试集。

## 5 实验与分析

### 5.1 实验平台

所有模型的训练环境和测试环境皆在 Ubuntu18.0.4 系统上搭建深度学习检测框架 Python3.8、Pytorch1.10, 硬件配置为 CPU Xeon-4210, GPU RTX3090。模型分类实验和检测实验均选取随机优化算法 (Adam) 优

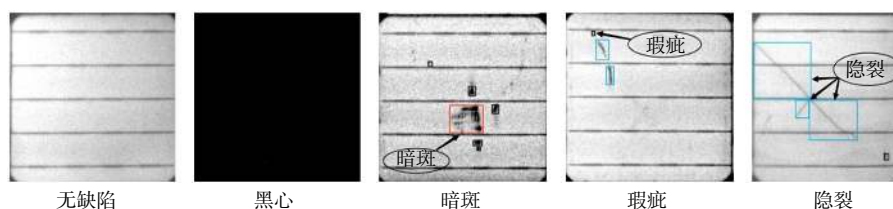


图 5 太阳能电池片种类

Fig. 5 Solar cell types

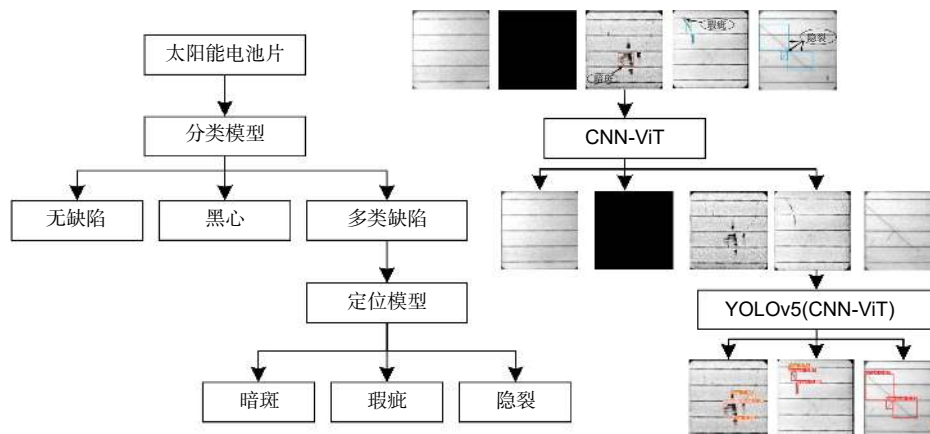


图 6 太阳能电池片缺陷检测流程

Fig. 6 Solar cell defect detection process



化网络参数, 初始学习率选取 0.01, 同时利用余弦 (Cosine) 函数调整学习率下降, 此外, 分类实验和检测实验不同参数如表 2 所示。

表 2 分类实验和检测实验不同参数值

Table 2 Classify and detect experimentally different parameter values

名称	分类实验	检测实验
输入图像分辨率	224 × 224	640 × 640
训练轮数 (epoch)	100	300
批量尺寸 (Batch size)	40	8

### 5.2 分类实验分析

为验证 CViT-Net 模型在太阳能电池片缺陷分类上的优势, 与先进卷积神经网络 ResNet50、DenseNet121、EfficientNet-B0、RegNet、MobileVit、MobileNetV3、ShuffleNetV2 和 GhostNet 等对比实验如表 3 和图 6 所示, 所有模型未采用预训练参数。模型对电池片测试采用精确度 (Precision)、召回率

(Recall) 和准确率 (Accuracy) 等评价指标, 它们的计算公式分别为:

$$Precision = \frac{Y_{TP}}{Y_{(TP+FP)}} \quad (9)$$

$$Recall = \frac{Y_{TP}}{Y_{(TP+FN)}} \quad (10)$$

$$Accuracy = \frac{Y_{(TP+TN)}}{Y_{(TP+FN+FP+FN)}} \quad (11)$$

其中:  $Y_{TP}$  表示正类样本被模型正确识别为正类的数量;  $Y_{TP+FP}$  表示模型预测的正类样本数量;  $Y_{TP+FN}$  表示真实的正类样本数量;  $Y_{TP+FN+FP+FN}$  表示总样本数量, 此外, 模型参数量 (Parameter) 和计算量 (FLOPs) 亦至关重要, 它们共同反映了模型复杂度。

由表 3 和图 7 的实验结果可知, 相较同参数量和计算量的经典卷积神经网络, CViT-Net-S 和 CViT-Net-L 模型分别实现了最优准确率 94.50% 和 95.10%, 模型参数量分别为 5.6 M 和 21.9 M, 模型计算量分别为 1.52 G 和 6.49 G。与专注于移动网络的 MobileVit

表 3 先进卷积神经网络算法对比

Table 3 Comparison of advanced convolutional neural network algorithms

模型	测试分辨率	Precision/%	Recall/%	Accuracy/%	Parameter/M	FLOPs/G
ResNet50	224×224	93.07	95.06	94.38	23.5	4.11
DenseNet121	224×224	90.03	93.78	91.60	6.9	2.86
EfficientNet-B0	224×224	92.33	94.11	93.60	5.3	0.39
RegNet	224×224	93.25	95.17	94.40	24.5	6.52
MobileVit	224×224	91.48	95.28	93.10	5.6	1.44
MobileNetV3	224×224	92.78	93.78	92.20	5.4	0.23
ShuffleNetV2	224×224	91.42	94.78	93.00	3.4	0.13
GhostNet	224×224	91.56	95.46	93.20	3.9	2.45
CViT-Net-S	224×224	93.00	95.94	94.50	5.6	1.52
CViT-Net-L	224×224	93.70	96.28	95.10	21.9	6.49

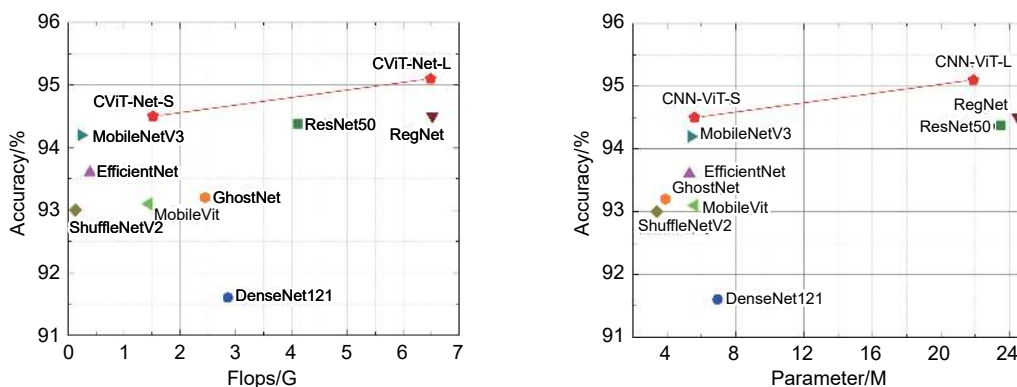


图 7 模型准确率相较计算量和参数量对比图

Fig. 7 Comparison chart of model accuracy compared to calculation amount and parameter amount

和 MobileNetV3 相比, CViT-Net-S 网络参数量和计算量表现不友好的原因是模型融入了 Transformer, 但在缺陷识别精确度指标上显著提升了 1.52% 和 0.22%, 召回率提升了 0.66% 和 2.16%。ShuffleNetV2 模型 FLOPs 为 0.13 G, 但相比于 CViT-Net-S 模型缺陷检测精确度、准确度和召回率显著提升了 1.8%、1.5% 和 1.16%。

CViT-Net 具有高精确、高召回率和轻量特点。CViT-Net-L 精确度和召回率高达 93.70% 和 96.28%, 相较于 DenseNet121, 精确度、召回率和准确度分别提升了 2.97%、2.16% 和 2.9%, 参数量和计算量分别减少 43.8% 和 46.8%; CViT-Net-L 与参数量相近的 RegNet 相比, 精确度、召回率和准确度分别提升了 0.45%、1.11% 和 0.7%, 参数量和计算量分别降低了 2.6 M 和 0.03 G; CViT-Net-S 与 DenseNet121、EfficientNet、ShuffleNetV2、GhostNet 等轻量级网络相比, 最优精确度、召回率和准确度高达 93%、95.96% 和 94.5%。参数量和计算量分别为 5.6 M 和 1.52 G。综上所述, CViT-Net 网络模型在低参数量和计算量的同时具有优秀的缺陷识别性能。

### 5.3 不同注意力应用于 CViT-Net 的性能比较

为验证 CA 注意力机制的有效性, 在保证其它模型参数量一致的情况下, 分别将 SE、CBAM、EMA 和 CA 注意力机制嵌入到 CViT-Net 模型中, 实验结果如表 4 所示。

实验结果表明, CViT-Net 模型在融入 CA 注意

力的情况下, 取得了最优的精确度、召回率和准确度, 分别达到 93%、95.94% 和 94.5%, 其参数量和计算量分别为 5.64 M 和 1.52 G。虽然 CA 注意力的参数量略高于 SE 和 CBAM, 但在精确度、召回率和准确度方面均取得了显著提升, 较 SE 分别增加了 12.26%、15.36% 和 6.82%, 较 CBAM 分别增加了 6%、8.34% 和 2.26%。相较参数量相同的 EMA, 精确度、召回率和准确度分别提升了 5.1%、9.54%、1.3%, 计算量降低了 0.2 M。综上所述, 将 CA 嵌入 CViT-Net 模型显著提升了检测效果。

### 5.4 消融实验

为验证 CViT-Net 模型在太阳能电池片缺陷检测的优越性, 分别对 G-C2F 模块、G-ViT 模块、CA 模块进行消融实验。为满足工业检测需要的高精度和高效率, 在 CViT-Net 系列中选择 CViT-Net-S 网络所做消融实验如表 5 所示。Baseline 表示卷积层和全连接层组合的缺陷识别方法。

“-”表示没有使用该模块, “√”表示使用该模块, 加粗字体表示最优项。实验结果表明, 基于卷积层和全连接层作为 Baseline 依次改进的 CViT-Net-S 网络在保持低参数量、低计算量的同时实现最优识别准确率、参数量和计算量分别为 94.50%、5.6 M 和 1.52 G。G-C2F 模块旨在提取电池片缺陷局部纹理细粒度特征, 相较原始 Baseline 基线在准确率上提升了 5.37%, 参数量和计算量分别减少了 39.8% 和 67.8%; G-ViT 模块是用于融合电池片缺陷局部特征和全局特征, 由

表 4 注意力机制性能比较

Table 4 Attention mechanism performance comparison

模型	Precision/%	Recall/%	Accuracy/%	Parameter/M	FLOPs/G
CViT-Net	79.86	80.25	86.78	4.59	4.2
CViT-Net+SE	80.74	80.58	87.68	5.60	1.53
CViT-Net+CBAM	87.0	87.6	92.24	5.63	1.54
CViT-Net+EMA	87.9	86.4	93.20	5.64	1.72
CViT-Net+CA	93.00	95.94	94.50	5.64	1.52

表 5 CViT-Net-S 网络消融实验

Table 5 CViT-Net-S network ablation experiment

模型	G-C2F	G-ViT	CA	Parameter/M	FLOPs/G	Accuracy/%
Baseline	-	-	-	14.8	4.2	86.78
CViT-Net-S	√	-	-	8.9	1.35	92.15
CViT-Net-S	√	√	-	12.6	1.51	92.31
CViT-Net-S	√	√	√	<b>5.6</b>	<b>1.52</b>	<b>94.50</b>

于 Transformer 融入使得参数量提升至 12.6 M, 但丰富的缺陷纹理特征促进 CViT-Net-S 在缺陷检测准确率高达 92.31%; 通过 CA 注意力强调缺陷特征, 抑制不相干的背景特征, 使得缺陷检测准确率达到了 93%; 为平衡参数量, CViT-Net-S 最终相较 Baseline 的参数量和计算量分别减少了 62.1% 和 63.8%, 准确率提升了 7.72%, 参数量降低了 9.2 M。

### 5.5 检测实验分析

由表 6 可知, 一、二阶段模型中 YOLOv5 在分类实验表现最优, mAP50 最高为 89.4%, 为验证

CViT-Net 网络在电池片缺陷检测具有优越性, 选择 YOLOv5 作为基础检测框架, 更换不同骨干网络进行实验验证。结果如表 7 所示, 采用 Precision、Recall 和平均精度均值 (mean average precision, mAP) 作为评价指标。其中 mAP50 是真实目标框面积与预测目标框面积的交集与并集之比大于 50% 的平均精度均值; mAP 是真实目标框面积与预测目标框面积的交集与并集之比大于 50% 且小于 95% 的平均精度均值, 它们有相同计算公式如式 (12) 所示,  $m$  表示缺陷类别数,  $AP_i$  表示第  $i$  个类别的平均精度。

表 6 不同目标检测算法实验对比

Table 6 Experimental comparison of different target detection algorithms

模型	mAP/%			mAP50/%
	隐裂	暗斑	瑕疵	
<b>Two stage:</b>				
Faster R-CNN( ResNet50)	86.1	85.4	76.8	82.8
Cascade-R-CNN( ResNet50)	89.3	86.8	79.9	85
Sparse R-CNN( ResNet50)	74.5	75.4	64.1	71.3
FoveaBox( ResNet50)	88.3	85.2	61.8	78.5
<b>One stage:</b>				
RetinaNet( ResNet50)	78.7	84.2	63.3	75.4
VFNet( ResNet50)	53.2	56.2	49.7	53
YOLOv5S	86.2	89.6	86.7	89.4
YOLOv6S	87.4	89.4	86.5	87.8
YOLOv7	80.8	86.7	81.2	82.9
YOLOv8S	86.9	88.1	86.0	87.0
YOLOX-S	88.1	88.8	87.2	88
PPYOLOE-S	87.7	89.9	84.5	87.4
YOLOv5(CViT-Net-S)	89.4	93.5	87.4	90.1
YOLOv5(CViT-Net-L)	<b>89.5</b>	<b>93.6</b>	<b>87.5</b>	<b>90.2</b>

表 7 YOLOv5 检测框架下的骨干网络对比实验

Table 7 YOLOv5 backbone network comparison experiment

检测框架	骨干网络	Precision/%	Recall/%	mAP/%	mAP50/%
YOLOv5	ResNet50	83.9	85.1	48.4	86.3
YOLOv5	DenseNet121	83.2	83.9	49.0	88.1
YOLOv5	EfficientNet	87.0	84.9	52.1	89.3
YOLOv5	RegNet	86.3	85.5	52.8	88.9
YOLOv5	MobileVit	82.9	82.0	49.9	87.4
YOLOv5	MobileNetV3	89.5	86.5	58.3	89.8
YOLOv5	ShuffleNetV2	83.7	80.4	48.6	86.8
YOLOv5	GhostNet	85.6	86.3	52.8	89.3
YOLOv5	CViT-Net-S	87.2	87.1	56.2	90.1
YOLOv5	CViT-Net-L	90.1	87.3	61.1	90.2

$$mAP = \frac{1}{m} \sum_{i=1}^m AP_i. \quad (12)$$

表中所有模型未采用骨干网络预训练参数和 YOLOv5 颈部及头部预训练参数, 初始学习率为 0.01, 使用 Adam 优化器进行 300 轮训练, 数据集采用多类缺陷图像 (隐裂、暗斑、瑕疵) 数据集, 共计 4970 片缺陷图像。

实验结果由表 6 可知, YOLOv5 检测框架, 骨干网络采用 CViT-Net-S 和 CViT-Net-L 实现的最优 mAP50 值分别是 90.1% 和 90.2%, mAP 值是 56.2% 和 61.1%。与 MobileVit、MobileNetV3、ShuffleNetV2 等骨干网络相比, CViT-Net-S 在召回率和 mAP 均达到了最优效果; 与 ResNet50、DenseNet121、RegNet 等网络相比, CViT-Net-S 和 CViT-Net-L 骨干网络在精确度、召回率、mAP 和 mAP50 值上保持最优, 其中提升效果明显的是 CViT-Net-L 骨干网络相较 ResNet50 在精确度和召回率上分别提升了 6.2% 和 2.2%, mAP 和 mAP50 上分别提升了 12.7% 和 3.9%。同时所有骨干网络在 YOLOv5 检测框架下的可视化定位结果如图 8 所示, CViT-Net 网络在保持隐裂、暗斑、瑕疵缺陷高定位精度时可避免模型漏检和误检, 保证了模型对缺陷的定位准确度。综合而言, 采用 G-C2F 模块提取电池片缺陷局部细粒度特征、G-ViT 模块融合电池片缺陷局部特征和全局特征的 CViT-Net

模型, 该模型作为骨干网络能充分提取电池片表面缺陷特征。

## 5.6 不同目标检测算法实验对比

为验证 CViT-Net-S 和 CViT-Net-L 作为骨干网络融合 YOLOv5 算法在太阳能电池片缺陷检测中具有高效性和泛化性, 与当前流行的一阶段、二阶段网络进行对比实验, 不同目标检测算法实验对比结果如表 7 所示。在太阳能电池片多类缺陷图像 (隐裂、暗斑、瑕疵) 数据集上进行对比实验。模型在没有预训练参数情况下重新训练 300 个 epoch, 图像训练和测试尺寸皆是 640 pixels×640 pixels, Faster RCNN (ResNet-50) 表示采用 Faster RCNN 为主体框架, ResNet-50 为骨干网络。

由表 6 可知, YOLOv5(CViT-Net)、YOLOv5(CViT-Net-L) 模型与一阶段、二阶段网络模型相比, 最优 mAP50 分别为 90.1% 和 90.2%, 利用 G-C2F 模块和 G-ViT 模块组合的特征提取网络 CViT-Net-S、CViT-Net-L, 在融合 YOLOv5 检测框架后可高效进行太阳能电池片表面缺陷检测, 相较经典 YOLO 和 R-CNN 系列算法在精确度、召回率、mAP 和 mAP50 等多个评价指标上皆占据优势。

## 6 结论

本文基于深度学习先进技术, 提出了一种能够高

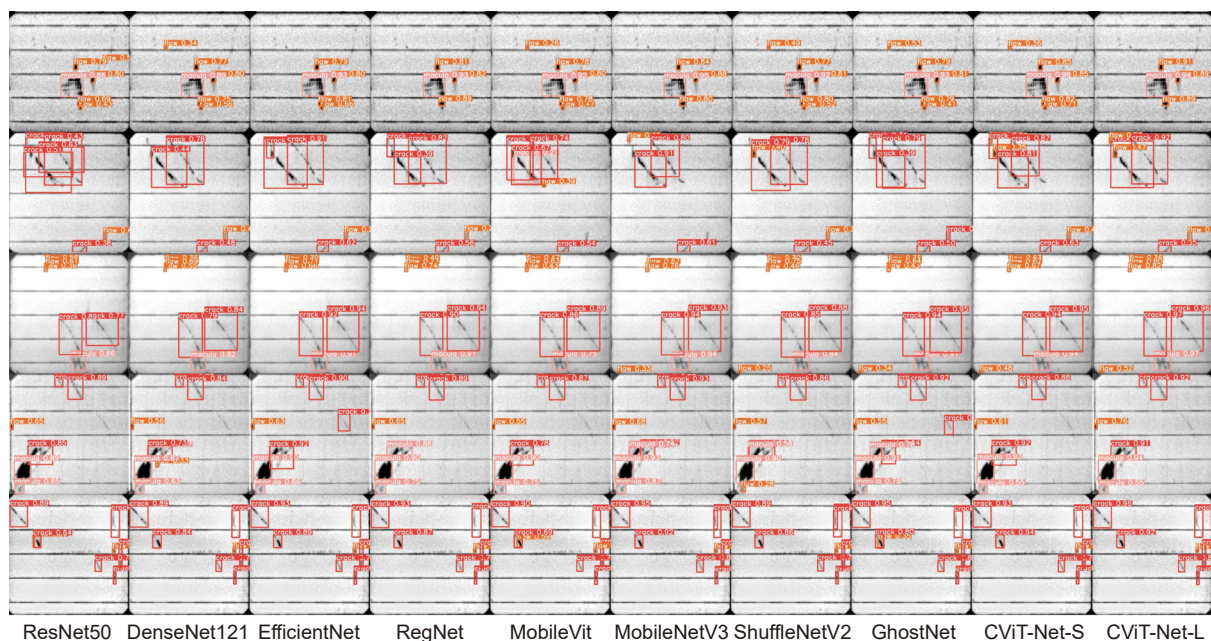


图 8 YOLOv5 检测框架下的可视化定位结果

Fig. 8 Visual positioning results under YOLOv5 detection framework

质量识别太阳能电池片表面缺陷的 CNN-ViT 模型。相较于经典的轻量级网络分类模型, CNN-ViT 模型在电池片缺陷分类精确度、召回率和准确度上都表现更优。同时, 为满足工业检测需求, 分别提出了参数量为 5.6 M 的 CViT-Net-S 结构和参数量为 21.9 M 的 CViT-Net-L 结构。此外, 采用 CViT-Net 作为骨干网络, 结合 YOLOv5 检测框架对太阳能电池片缺陷进行定位。与经典的轻量级骨干网络或先进的 YOLO 系列模型相比, CViT-Net 在结合 YOLOv5 检测框架后实现了出色的缺陷检测精度和缺陷定位精度。在后续工作中, 将深入拓展 CViT-Net 模型, 以应用于其他实物的缺陷分类检测, 满足不同实物缺陷识别的需求。

## 参考文献

- [1] Yu X, Yan J S, Wu Z P, et al. Research progress of solar desalination materials produced by laser micro-nano fabrication[J]. *Opto-Electron Eng*, 2022, **49**(1): 210313.  
余星, 严俊森, 吴志鹏, 等. 激光微纳制造太阳能海水淡化材料研究进展[J]. *光电工程*, 2022, **49**(1): 210313.
- [2] Herraiz Á H, Marugán A P, Márquez F P G. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure[J]. *Renew Energy*, 2020, **153**: 334–348.
- [3] Zeng D Y, Liang Z X, Wu Z Z. Optimal mean linear classifier via weighted nuclear norm and  $L_{2,1}$  norm[J]. *J Electron Inf Technol*, 2022, **44**(5): 1602–1609.  
曾德宇, 梁泽道, 吴宗泽. 基于加权核范数和  $L_{2,1}$  范数的最优均值线性分类器[J]. *电子与信息学报*, 2022, **44**(5): 1602–1609.
- [4] Juan R O S, Kim J. Photovoltaic cell defect detection model based-on extracted electroluminescence images using SVM classifier[C]//2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC), Fukuoka, Japan, 2020: 578–582. <https://doi.org/10.1109/ICAIC48513.2020.9065065>.
- [5] Wang C, Jiang X Y, Liu X H. Defect detection in crystalline silicon solar cells based on electroluminescence imaging[J]. *J Optoelectron-Laser*, 2011, **22**(9): 1332–1336.  
王超, 蒋晓瑜, 柳效辉. 基于电致发光成像理论的硅太阳能电池缺陷检测[J]. *光电子·激光*, 2011, **22**(9): 1332–1336.
- [6] Firuzi K, Vakilian M, Phung B T, et al. Partial discharges pattern recognition of transformer defect model by LBP & HOG features[J]. *IEEE Trans Power Delivery*, 2019, **34**(2): 542–550.
- [7] Juan R O S, Kim J. Photovoltaic cell defect detection model based-on extracted electroluminescence images using SVM classifier[C]//2020 International Conference on Artificial Intelligence in Information and Communication (ICAIC). IEEE, 2020: 578–582. <https://doi.org/10.1109/ICAIC48513.2020.9065065>.
- [8] Li Y, Li Y J, Liu J C, et al. Research on segmentation of steel surface defect images based on improved Res-UNet network[J]. *J Electron Inf Technol*, 2022, **44**(5): 1513–1520.  
李原, 李燕君, 刘进超, 等. 基于改进 Res-UNet 网络的钢铁表面缺陷图像分割研究[J]. *电子与信息学报*, 2022, **44**(5): 1513–1520.
- [9] Masita K, Hasan A, Shongwe T. 75MW AC PV module field anomaly detection using drone-based IR orthogonal images with Res-CNN3 detector[J]. *IEEE Access*, 2022, **10**: 83711–83722.
- [10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016: 770–778. <https://doi.org/10.1109/CVPR.2016.90>.
- [11] Du B L, He Y G, He Y Z, et al. Intelligent classification of silicon photovoltaic cell defects based on eddy current thermography and convolution neural network[J]. *IEEE Trans Industr Inform*, 2020, **16**(10): 6242–6251.
- [12] Chen H Y, Pang Y, Hu Q D, et al. Solar cell surface defect inspection based on multispectral convolutional neural network[J]. *J Intell Manuf*, 2020, **31**(2): 453–468.
- [13] Su B Y, Chen H Y, Zhou Z. BAF-detector: an efficient CNN-based detector for photovoltaic cell defect detection[J]. *IEEE Trans Ind Electron*, 2022, **69**(3): 3161–3171.
- [14] 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.  
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.
- [15] Zhang M, Yin L J. Solar cell surface defect detection based on improved YOLO v5[J]. *IEEE Access*, 2022, **10**: 80804–80815.
- [16] Chen X, Peng D L, Gu Y. Real-time object detection for UAV images based on improved YOLOv5s[J]. *Opto-Electron Eng*, 2022, **49**(3): 210372.  
陈旭, 彭冬亮, 谷雨. 基于改进 YOLOv5s 的无人机图像实时目标检测[J]. *光电工程*, 2022, **49**(3): 210372.
- [17] Howard A, Sandler M, Chen B, et al. Searching for MobileNetV3[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019: 1314–1324. <https://doi.org/10.1109/ICCV.2019.00140>.
- [18] Zhang X Y, Zhou X Y, Lin M X, et al. ShuffleNet: an extremely efficient convolutional neural network for mobile devices[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 6848–6856. <https://doi.org/10.1109/CVPR.2018.00716>.
- [19] Koonce B. EfficientNet[M]//Koonce B. *Convolutional Neural Networks with Swift for Tensorflow: Image Recognition and Dataset Categorization*. Berkeley: Apress, 2021: 109–123. [https://doi.org/10.1007/978-1-4842-6168-2\\_10](https://doi.org/10.1007/978-1-4842-6168-2_10).
- [20] Han K, Wang Y H, Tian Q, et al. GhostNet: more features from cheap operations[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020: 1577–1586. <https://doi.org/10.1109/CVPR42600.2020.00165>.
- [21] Radosavovic I, Kosaraju R P, Girshick R, et al. Designing network design spaces[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, USA, 2020: 10425–10433. <https://doi.org/10.1109/CVPR42600.2020.01044>.
- [22] Han K, Xiao A, Wu E H, et al. Transformer in transformer[C]//35th Conference on Neural Information Processing Systems, 2021: 15908–15919.
- [23] Wu H P, Xiao B, Codella N, et al. CvT: introducing convolutions to vision transformers[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, Canada, 2021: 22–31. <https://doi.org/10.1109/ICCV48922.2021.00009>.

- [24] d'Ascoli S, Touvron H, Leavitt M L, et al. ConViT: improving vision transformers with soft convolutional inductive biases[J]. *J Stat Mech Theory Exp*, 2022, **2022**(11): 114005.
- [25] Mehta S, Rastegari M. MobileViT: light-weight, general-purpose, and mobile-friendly vision transformer[C]//*Tenth International Conference on Learning Representations*, 2022. .
- [26] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Vancouver, 2023: 7464–7475. <https://doi.org/10.1109/CVPR52729.2023.00721>.
- [27] Pawar N, Waghmare A, Pratap A, et al. Miniscule object detection in aerial images using YOLOR: a review[M]//Kumar S, Hiranwal S, Purohit S D, et al. *Proceedings of International Conference on Communication and Computational Technologies: ICCCT 2022*. Singapore: Springer, 2023: 697–708. [https://doi.org/10.1007/978-981-19-3951-8\\_52](https://doi.org/10.1007/978-981-19-3951-8_52).
- [28] Norkobil Saydirasulovich S, Abdusalomov A, Jamil M K, et al. A YOLOv6-based improved fire detection approach for smart city environments[J]. *Sensors*, 2023, **23**(6): 3161.
- [29] Talaat F M, Zaineldin H. An improved fire detection approach based on YOLO-v8 for smart cities[J]. *Neural Comput Appl*, 2023, **35**(28): 20939–20954.
- [30] Zhang Y C, Zhang W B, Yu J Y, et al. Complete and accurate holly fruits counting using YOLOX object detection[J]. *Comput Electron Agric*, 2022, **198**: 107062.
- [31] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, 2017: 2261–2269. <https://doi.org/10.1109/CVPR.2017.243>.
- [32] Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//*2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, USA, 2017: 936–944. <https://doi.org/10.1109/CVPR.2017.106>.
- [33] Liu S, Qi L, Qin H F, et al. Path aggregation network for instance segmentation[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018: 8759–8768. <https://doi.org/10.1109/CVPR.2018.00913>.
- [34] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, USA, 2021: 13708–13717. <https://doi.org/10.1109/CVPR46437.2021.01350>.
- [35] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, 2018: 7132–7141. <https://doi.org/10.1109/CVPR.2018.00745>.
- [36] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//*Proceedings of the 15th European Conference on Computer Vision (ECCV)*, Munich, 2018: 3–19. [https://doi.org/10.1007/978-3-030-01234-2\\_1](https://doi.org/10.1007/978-3-030-01234-2_1).
- [37] Ouyang D L, He S, Zhang G Z, et al. Efficient multi-scale attention module with cross-spatial learning[C]//*ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 2023: 1–5. <https://doi.org/10.1109/ICASSP49357.2023.10096516>.
- [38] Szegedy C, Vanhoucke V, Ioffe S, et al. Rethinking the inception architecture for computer vision[C]//*2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, USA, 2016: 2818–2826. <https://doi.org/10.1109/CVPR.2016.308>.
- [39] Wu C C, Hung Y C. A study on electroluminescent phenomenon for solar cells test[C]//*2018 1st IEEE International Conference on Knowledge Innovation and Invention (ICKII)*, Jeju, Korea (South), 2018: 172–175. <https://doi.org/10.1109/ICKII.2018.8569147>.

## 作者简介



陶志勇 (1978-), 男, 博士, 教授, 主要研究方向为机器学习、图像处理、智能信息处理。  
E-mail: xyzmail@126.com



易廷军 (1997-), 男, 硕士研究生, 主要研究方向为深度学习目标检测。  
E-mail: 1185594256@qq.com



【通信作者】何燕 (1994-), 女, 硕士研究生, 主要研究方向为计算机视觉、图像处理。  
E-mail: 2575561115@qq.com



张尧晟 (1997-), 男, 硕士研究生, 主要研究方向为深度学习自然语言处理。  
E-mail: 2973303100@qq.com



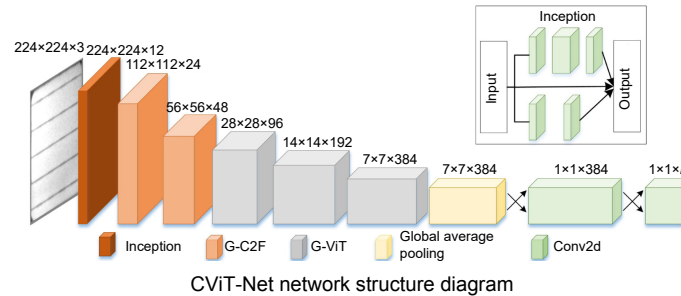
林森 (1980-), 男, 博士, 副教授, 主要研究方向为图像处理与模式识别。  
E-mail: lin\_sen6@126.com



扫描二维码, 获取PDF全文

# Surface defect detection of solar cells using local and global feature fusion

Tao Zhiyong<sup>1</sup>, He Yan<sup>1\*</sup>, Lin Sen<sup>2</sup>, Yi Tingjun<sup>1</sup>, Zhang Yaosheng<sup>1</sup>



**Overview:** The methods employed for identifying surface defects on solar cell panels encompass traditional machine learning and deep learning. Traditional machine learning methods have advantages in defect recognition and well-established algorithms for detecting surface defects on solar cell panels. However, these methods encounter challenges, including extensive parameter tuning, issues with model robustness, suboptimal generalization performance, and reliance on engineers' subjective experience for defect discrimination in solar cell defect detection. Moreover, they need help adapting to prolonged manual labor. In contrast, deep learning methods face challenges from the high similarity of defect features on solar cell panels and the complexity of background features. Issues such as insufficient extraction of fine-grained defect features and feature loss during network deepening may arise, resulting in decreased detection accuracy. The surface defects on solar cell panels show significant intra-class and minimal inter-class differences, combined with a complex background. Therefore, achieving high-precision automatic detection of surface defects on solar cell panels becomes challenging. We utilize advanced techniques in deep learning and computer vision to address this issue. We propose a method named Convolutional-Vision Transformer Networks (CViT-Net), specifically designed to efficiently integrate local and global features for accurate defect detection in solar cell panels. The model initially utilizes a Ghost Focus (G-C2F) module to extract local features related to defects in solar cell panels. Subsequently, a coordinate attention mechanism is introduced to emphasize defect features and attenuate background features. Finally, we construct a Ghost Vision (G-ViT) module to integrate local and global features of defects in solar cell panels. To address various demands for detection accuracy and model parameterization, we introduce the CViT-Net-S structure with a parameter count of 5.6 M and the CViT-Net-L structure with a parameter count of 21.9 M, serving diverse practical applications in low-resource and high-resource environments, respectively. Experimental results illustrate the remarkable performance of our model in classifying and detecting defects in solar cell panels. Compared to lightweight models like MobileVit, MobileNetV3, and GhostNet, our CViT-Net-S model achieves accuracy improvements of 1.4%, 2.3%, and 1.3%, respectively, for defect classification in solar cell panels and mAP50 enhancements of 2.7%, 0.3%, and 0.8%, respectively, in defect detection. Compared to RecNet50 and RegNet, the CNN-ViT-L model demonstrates classification accuracy enhancements of 0.72% and 0.7% and mAP50 improvements of 3.9% and 1.3%, respectively. When compared to advanced object detection models like YOLOv6, YOLOv7, and YOLOv8, CViT-Net-S and CViT-Net-L, serving as backbone networks, continue to demonstrate robust detection performance in terms of mAP and mAP50 metrics. These results underscore the algorithm's significant practical value in the surface defect detection field of solar cell panels. In future work, we plan to extend the CViT-Net model for application in defect classification detection for other physical entities to meet diverse defect recognition needs.

Tao Z Y, He Y, Lin S, et al. Surface defect detection of solar cells using local and global feature fusion[J]. *Opto-Electron Eng*, 2024, 51(1): 230292; DOI: 10.12086/oe.2024.230292

Foundation item: Project supported by Applied Basic Research Projects of Department of Science & Technology of Liaoning Province(2022JH2/101300274), Research Project on Teaching Reform of Graduate Education in Liaoning Province (LNYJG2023117), and Basic Research Project of Liaoning Provincial Department of Education (LJKMZ20220676)

<sup>1</sup>School of Electronic and Information Engineering, Liaoning Technical University, Huludao, Liaoning 125105, China; <sup>2</sup>School of Automation and Electrical Engineering, Shenyang Ligong University, Shenyang, Liaoning 110159, China

\* E-mail: 2575561115@qq.com