

全局与局部多尺度特征融合晶圆缺陷分类网络^{*}

陈晓雷 李正成 杨富龙 温润玉 沈星阳

(兰州理工大学电气工程与信息工程学院 兰州 730000)

摘要:在半导体制造领域,晶圆缺陷分类是确保产品质量的重要步骤。然而,由于晶圆缺陷的多样性和复杂性,现有的混合型晶圆缺陷分类网络在准确性上仍然存在不足。针对这一问题,提出了一种基于全局和局部多尺度特征融合的混合型晶圆缺陷分类网络—MLG-Net。MLG-Net 由 3 个主要模块组成:特征提取模块、全局分支和局部分支。该网络旨在更好地提取和利用晶圆缺陷图像的全局语义信息与局部细节特征,这两种特征通过多尺度特征融合技术相结合,最终形成一个更加全面的特征表示,有助于分类器在面对复杂混合缺陷时,做出更为准确的判断,从而提升分类精度。为了验证 MLG-Net 的有效性,在包含 38 种混合类型缺陷的数据集—MixedWM38 上进行了大量实验,其分类准确度达到 98.84%。结果表明,MLG-Net 在综合性能上优于当前主流的六种晶圆缺陷分类方法。这一结果证明了全局与局部特征融合在处理混合型晶圆缺陷分类任务中的重要性和有效性。

关键词:半导体制造;晶圆缺陷分类;混合型缺陷;全局与局部特征;特征融合

中图分类号: TN305;TP18;TP391.41 **文献标识码:** A **国家标准学科分类代码:** 510.4050

Wafer defect classification network with global and local multi-scale feature fusion

Chen Xiaolei Li Zhengcheng Yang Fulong Wen Runyu Shen Xingyang

(College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou 730000, China)

Abstract: In semiconductor manufacturing, wafer defect classification is an important step in ensuring product quality. However, due to the diversity and complexity of wafer defects, the existing hybrid wafer defect classification network still has shortcomings in accuracy. To solve this problem, a hybrid wafer defect classification network based on global and local multi-scale feature fusion—MLG-Net was proposed. MLG-Net consists of three main modules: feature extraction module, global branch, and local branch. The network aims to better extract and utilize the global semantic information and local detail features of wafer defect images, which are combined with multi-scale feature fusion technology to form a more comprehensive feature representation, which helps the classifier to make more accurate judgments in the face of complex mixed defects, thereby improving the classification accuracy. To verify the effectiveness of MLG-Net, a large number of experiments were carried out on MixedWM38, a dataset containing 38 mixed types of defects, and the classification accuracy reached 98.84%. The results show that MLG-Net is superior to the six mainstream wafer defect classification methods in terms of comprehensive performance. This result demonstrates the importance and effectiveness of global and local feature fusion in dealing with hybrid wafer defect classification tasks.

Keywords: semiconductor manufacturing; wafer defect classification; mixed defects; global and local features; feature fusion

0 引言

近年来,随着 5G 网络、物联网(internet of things,

IoT)、无人驾驶等新兴技术的发展,半导体芯片的市场需求日益增大。晶圆(wafer)是制造半导体芯片的基础,晶圆图(wafer map)详细展示了半导体芯片制造过程中晶圆上的单个裸片(die)状态。通过分析晶圆缺陷图像,能

够查明晶圆缺陷产生的原因,从而进行消除并提高产品良率^[1]。传统晶圆缺陷识别由经验丰富的工人来完成,但这种依靠人工目视来检测的方法存在成本高昂、效率低下且主观性强等诸多缺点。伴随着集成电路层数和晶圆上芯片数量的急剧增加,电路的规模和数量都使得传统晶圆缺陷分类方法难以完成检测任务^[2]。此外,由于晶圆生产过程中会产生对人体有害的化学物质,因此长期的人工检测可能会对工人身体造成伤害。姜也等^[3]提出一种针对可编程门阵列(field programmable gate arrays, FPGA)单板双倍数据速率(double data rate, DDR)存储芯片,建立了缺陷检测模型并开展了芯片内部焊点故障的检测研究方法。史浩琛等^[4]提出了一种基于改进 YOLOv4 算法的高精度晶圆缺陷检测方法,利用 CCD 工业相机和高倍率光学显微镜采集图像。因此,亟需利用深度学习等人工智能技术实现高效的晶圆缺陷分类。

现有晶圆缺陷分类方法主要针对单一类型缺陷^[5-7],即晶圆上只存在一种缺陷,且获得了较高的分类准确性。然而,随着设计和制造复杂性的增加,晶圆上出现了不同类型的缺陷,从而产生混合型缺陷^[8]。目前针对混合型晶圆缺陷的分类方法还较少,代表性工作如下:Tello 等^[9]提出了一种新的深度结构化机器学习方法,扩展了随机通用回归网络模型,用于识别和分类晶圆上的单一缺陷和混合缺陷模式。Wang 等^[10]提出了一种可变形卷积网络(deformable convolutional networks, DC-Net),并且采用独热编码机制来进行混合晶圆缺陷分类。Nag 等^[11]提出一个名为 WaferSegClassNet(WSCN)的网络,通过采用编码器-解码器架构,WSCN 能够对混合型晶圆缺陷进行分类和分割。Wei 等^[12]提出了一种基于多分辨率小波集成注意力网络(multiresolution wavelet integrated attention network, MRWA-Net)的晶圆缺陷识别方法。通过离散小波变换(discrete wavelet transform, DWT)层,该方法扩展了卷积神经网络(convolutional neural networks, CNN)的特征学习空间,有助于从不同频率成分和位置信息中提取隐藏信息。Luo 等^[13]设计了一种名为(composite wafer defect recognition framework, CWDR-Net)的复合晶圆缺陷识别框架,旨在解决制造过程中可能出现的复杂缺陷类型。该框架利用了多视图动态特征增强(multiview dynamic feature enhancement, MVDFE)模块和特定类别分类器,能够从缺陷模式中选择性地提取信息,并为每种基本缺陷类型进行特定类别的识别。虽然上述工作都有不错的表现,但仍然存在以下问题:

1) 在晶圆的制造过程中,缺陷的种类、尺寸、大小等都是不确定的,并且交叉重叠,这使得特征更加隐蔽,增加了错误识别或者漏识别的可能性。

2) 通常,普通的 CNN 会增加通道数量以获取更多信息。但这可能会将一些不太有用的信息引入到不同的通

道中,而且每个通道提供的信息也可能比较零散,需要进行整合。

针对现有方法的缺陷,提出了一种基于全局和局部多尺度特征融合的混合型晶圆缺陷分类网络 MLG-Net。该网络由特征提取模块、多尺度通道注意力模块、全局分支和局部分支构成,实验证明在包含 38 种混合类型缺陷的 MixedWM38 数据集上,MLG-Net 综合性能优于 6 种先进的晶圆缺陷分类方法。

1 相关工作

1.1 单一类型晶圆缺陷分类方法

近年来,机器学习和深度学习的方法在半导体制造中晶圆图图案分类的自动化中得到了有效的应用。Kang 等^[14]提出了一种混合方法来改善半导体制造中晶圆图图案分类的自动化。首先,构建了两个基本分类器,分别利用传统的手工特征提取和卷积神经网络分类的方法。然后,通过堆叠集成将它们的输出组合在一起,以进行最终预测。传统的(deep convolutional neural networks, DCNN)严重依赖高质量的数据集进行训练。然而,在实践中很难获得平衡和充分的标记数据。Xu 等^[15]针对晶圆图检测中的数据集不平衡问题提出了一种深度学习方法。通过引入注意力机制和余弦归一化,该方法能够从不平衡数据集中学习到更多鲁棒的知识。Wang 等^[16]提出了一种基于变分自编码生成对抗网络(variational autoencoder generative adversarial network, VAE-GAN)的晶圆缺陷分类(wafer defect classification, WDC)方法,用于解决 WDC 中的数据不平衡问题。该方法结合了 VAE 和 GAN 的优势,生成接近真实样本的新图像,从而改善了 WDC 系统的性能。Tsai 等^[17]提出了一种名为全局到局部生成对抗网络(global-to-local generative adversarial network, G2LGAN)的方法,用于解决晶圆图缺陷分类中的数据不平衡和数据不足问题。该方法结合了深度学习框架,能够分别提取全局和局部特征,并使用随机欠采样处理数据不平衡。Doss 等^[18]结合了预训练的 ShuffleNet-v2 和 CNN 架构,提出了一种基于深度学习的硅片缺陷识别与分类模型,主要是利用晶圆图影像来辨识与分类晶圆的缺陷。Chen 等^[19]提出了一种新的卷积神经网络的方法,名为(wafer structural defect detection, WDD-Net)。用于快速准确地识别和分类晶圆结构缺陷。通过设计图像采集系统和数字图像预处理技术,将硅片图像分割成硅晶粒图像,并采用深度可分离卷积和全局平均池化减少参数和计算量。Saqlain 等^[20]提出了一种采用数据增强克服类不平衡问题,利用正则化等方法提高模型分类性能,用于晶圆图分类。Wang 等人^[21]提出了一种带有残差块的 DCNN 模型,在模型中加入平衡函数来提高模

型性能。Yoon 等^[22]提出了一种半自动晶圆图图案分类方法,该方法采用手动和自动的选项,给定一个阈值,如果不确定性低于某个阈值,则会被自动识别,否则,需要工艺工程师手动分类。Chen 等^[23]提出了一种双通道卷积神经网络来进行晶圆图像分类的方法,利用两种不同神经网络形成双通道来提取特征后得到新的多源特征,进一步对两个通道中新的特征源处理后,将新特征输入纠错码与支持向量机的组合模型中进行分类。

1.2 混合类型晶圆缺陷分类方法

为了识别混合型晶圆缺陷,Cha 等^[24]提出了一种改进的 (convolutional networks for biomedical image segmentation, U-Net) 网络模型,利用残差注意块结合注意机制和残差块来分割混合缺陷,以解决半导体中复杂缺陷检测的问题。Wei 等^[25]提出了一种将 Transformer 和 CNN 相结合的模型,用于识别具有混合类型的晶圆缺陷。Shim 等^[26]提出了一种训练 CNN 的方法。该方法通过混淆、随机旋转和噪声滤波 3 个组件合成晶圆图,用于训练 CNN 来分类混合缺陷晶圆图。Wang 等^[27]提出了一种知识增强广义学习系统,用于混合型晶圆缺陷检测。该系统包含知识模块和广义选择采样模块,利用多通道选择性采样网络解耦混合型缺陷。Byun 等^[28]提出了一种用于晶圆图合成的深度卷积生成对抗网络 (deep convolutional generative adversarial network for wafer map synthesis, DCGAN-WS), 该网络通过生成单一类型图案并添加像素求和来生成混合型缺陷,来解决模型缺乏标记数据的问题。

2 晶圆缺陷分类

2.1 整体框架

混合型晶圆缺陷存在多种多样的特征,包括缺陷的位置、角度、形状、大小和尺寸等。这些特征的多样性对于缺陷检测和分类具有挑战性。为了提高混合型晶圆缺陷的识别准确率,提出一种基于全局和局部多尺度特征融合的分类方法 MLG-Net。该方法的总体结构如图 1 所示,主要由特征提取模块 (feature extraction module, FEM)、全局分支 (global branch, GB) 和局部分支 (local branch, LB) 组成。首先利用特征提取模块分别提取输入晶圆图像的浅层和深层特征,然后这些特征分别被送入到全局分支、局部分支中。全局分支主要以 ResNet50 为骨干网络,通过一系列卷积和残差模块从输入特征中逐步提取出全局语义信息,这些特征对于后续晶圆缺陷分类至关重要。局部分支由多尺度通道注意力 (multi-scale channel attention, MSCA) 模块和特征增强自注意力 (features enhance self-attention FESA) 模块构成,多尺度通道注意力模块旨在获取不同层次的空间上下文信息,有助于模型更好的理解特征中的形状信息。特征增强自注意力模块主要处理特征中的局部细节信息,例如边缘和纹理等,这些细节对于晶圆缺陷的细粒度分类起决定性作用。最后来自全局分支、局部分支的不同尺度特征进行融合后被送入全局平均池化层和全连接层进行晶圆缺陷分类。

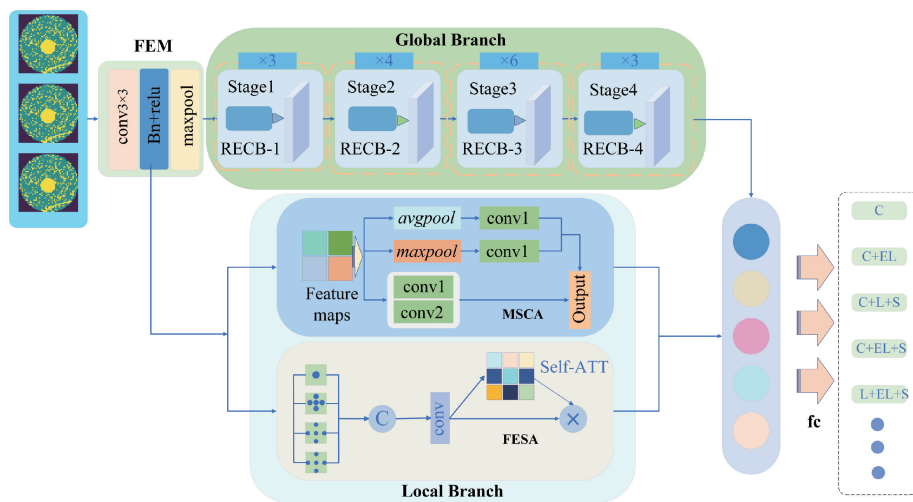


图 1 模型整体框架

Fig. 1 The overall framework of the model

2.2 全局分支

全局分支主要由 4 个 $\text{RECB-}i (i=1, 2, 3, 4)$ 模块构成,将轻量级卷积注意力模块 (convolutional block

attention module, CBAM)^[29]嵌入到 ResNet50 的 4 个残差学习块 ($\text{conv}2-3$, $\text{conv}3-4$, $\text{conv}4-6$, $\text{conv}5-3$) 中形成 RECB,其详细结构图如图 2 所示。

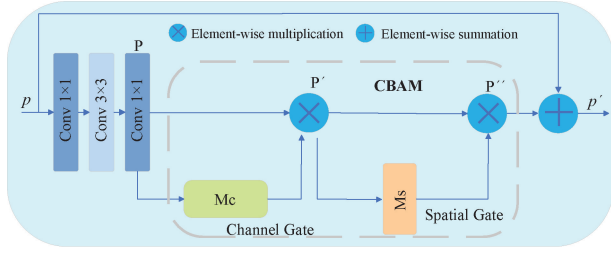


图 2 RECB 模块

Fig. 2 RECB module

RECB 的工作原理为:首先使用一个 1×1 卷积来降低输入特征图的维度,然后使用一个 3×3 卷积进行特征提取,最后再通过一个 1×1 卷积进行维度的升维,恢复到原始的维度。然后经过注意力机制 Mc 后得到注意力权重 $Mc(P)$, $Mc(P)$ 和 P 相乘后得到特征 P' ; P' 再经过一个空间注意力机制 Ms 得到注意力权重 $Ms(P')$, $Ms(P')$ 和 P' 相乘得到特征 P'' , P'' 与残差 p 相加后得到最终的输出 p' 。这样的设计可以显著减少中间特征图的维度,在降低计算复杂度的同时保持模型的性能,并且提高了计算效率。这个过程可以表述为:

$$P = \text{conv}_{1 \times 1}(\text{conv}_{3 \times 3}(\text{conv}_{1 \times 1}(p))) \quad (1)$$

$$P' = M_c(P) \otimes P \quad (2)$$

$$P'' = M_s(P') \otimes P' \quad (3)$$

$$p' = P'' + p \quad (4)$$

其中, $\otimes$ 表示对应元素相乘操作, $\oplus$ 表示对应元素相加操作。

2.3 局部分支

局部分支由 MSCA 模块和 FESA 模块组成。MSCA 模块的主要目的是获取不同层次的空间上下文信息,从而帮助模型更好地理解特征中的形状信息。FESA 模块则侧重于处理特征中的局部细节信息,例如边缘和纹理等,这些细节对于晶圆缺陷的细粒度分类至关重要。

1) 多尺度通道注意力模块 (MSCA)

混合型晶圆缺陷分类的难点在于缺陷变化的多样性和复杂性,因此提出了一种多尺度通道注意力模块来捕获特征中不同尺寸的特征信息,从而提高模型对不同缺陷的识别能力。如图 3 所示, MSCA 模块利用多个并行的卷积分支来处理输入特征张量,以捕获不同尺度下的特征信息。其中,一个分支经过一层 1×1 和 3×3 的卷积操作,以提取局部细节特征。而另一个分支经过一系列不同尺寸的卷积核(包括 1×3 、 3×1 、 1×5 、 5×1 、 1×7 和 7×1)来捕获更大范围的上下文信息。这个过程可以表述为:

$$F_1 = \text{conv}_{3 \times 3}(\text{conv}_{1 \times 1}(f)) \quad (5)$$

$$f' = \text{conv}_{1 \times 1}(f) \quad (6)$$

$$F_2 = \text{Aconv}_{3 \times 3}(f') + \text{Aconv}_{5 \times 5}(f') + \text{Aconv}_{7 \times 7}(f') \quad (7)$$

其中, f 为来自特征提取模块的输入特征图, $\text{conv}_{1 \times 1}$ 、 $\text{conv}_{3 \times 3}$ 分别代表 1×1 卷积、 3×3 卷积, $\text{Aconv}_{i \times i}$ 为非对称卷积,其中 i ($i=3, 5, 7$) 为卷积核大小。

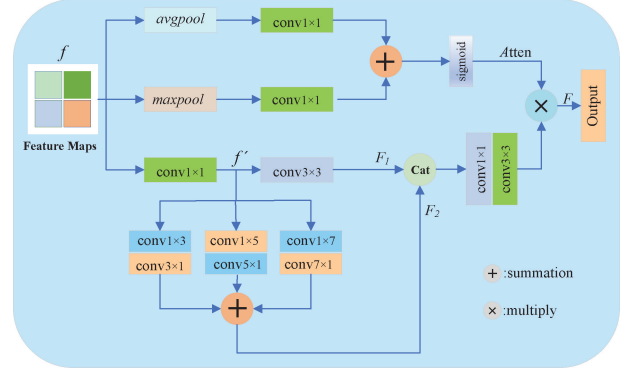


图 3 多尺度通道注意力模块

Fig. 3 multi-scale channel attention module

此外,还增加了一个通道加权支路。该支路通过全局平均池化和最大池化操作提取输入特征张量的全局信息,然后使用一个卷积层结合 sigmoid 函数生成注意力权重,用于动态调整特征图权重,从而突出重要特征信息。这个过程可以表述为:

$$\text{Atten} = \sigma(\text{conv}_{1 \times 1}(\text{avgpool}(f)) + \text{conv}_{1 \times 1}(\text{maxpool}(f))) \quad (8)$$

式中: σ 表示 sigmoid 函数, avgpool 和 maxpool 分别是空间维度的平均池化和最大池化。整个 MSCA 模块的过程可以描述为:

$$F = \text{Atten} \otimes \text{conv}_{3 \times 3}(\text{conv}_{1 \times 1}(\text{Cat}(F_1, F_2))) \quad (9)$$

式中: $\otimes$ 表示元素乘法, Cat 表示特征拼接, F 表示 MSCA 的输出特征图。

2) 特征增强自注意力模块 (FESA)

特征增强自注意力模块结构如图 4 所示,主要由膨胀卷积模块 Multi-Atrous 和自注意力块 Self-ATT 构成。

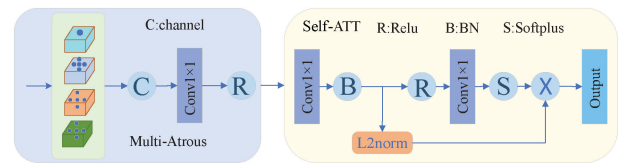


图 4 特征增强自注意力模块

Fig. 4 Feature-enhanced self-attention module

膨胀卷积模块包含 4 个膨胀卷积层,其目的是通过多个不同的膨胀率来扩大感受野,以捕获特征中的局部细节信息。在膨胀卷积模块初始化阶段创建了 4 个分支,每个分支使用不同尺寸 (1×1 和 3×3) 的卷积核和膨胀率 (2, 3, 4) 来实现不同尺度的特征提取,用来获得不同空间感受野和全局平均池化分布的特征图,多个特征图融合后经过 1×1 卷积和 ReLU 激活函数,将输出的特

征传送到自注意力块中。自注意力块由卷积层、BN、ReLU 和 Softplus 构成。首先使用 $cnov_{1\times 1}$ 和 BN 对特征进行处理,然后通过 1×1 卷积生成注意力图后进行 Softplus 操作,用于动态调整局部特征的重要性。最后,通过归一化 L2norm 和逐元素乘法操作,将局部特征与注意力图相乘,得到最终的注意力加权局部特征表示图来进一步提高模型对输入整体的理解能力。

3 混合型晶圆图数据集

使用 MixedWM38 数据集来评估所提出模型的有效性,MixedWM38 包含 38 个类别,分别为 8 种单一缺陷、13 种二重混合缺陷、12 种三重混合缺陷和 4 种四重混合缺陷,如图 5 中(a)、(b)、(c)、(d)所示。C1 为正常无缺陷类别,C2 到 C9 为常见的单一类型缺陷晶圆图。C2 为中

心缺陷,主要集中在晶圆图中心区域。C3 为“甜甜圈”(Donut)缺陷,主要集中在一个位于晶圆图中心附近的环状区域。C4 为 Edge-Loc 缺陷,主要聚集在晶圆图局部边缘区域附近。C5 为 Edge-Ring 缺陷,主要分布在晶圆图边缘。C6 为局部型缺陷,主要分布在晶圆图边缘附近,成簇状。C7 为 Near-full 缺陷,缺陷占据整个晶圆图区域。C8 为划痕缺陷,主要表现为弯曲形状的划痕。C9 被指定为随机缺陷,缺陷模式不规则。C10-C38 为混合型缺陷,由单一类型缺陷混合而成。例如,C38 由 D+L+ER+S 4 个单一缺陷混合而成。由于 C7Near-full 缺陷和 C9 随机缺陷容易识别,它们分别只有 866 张和 149 张图像,其余类别都为 1 000 张图像,总共 38 015 张。MixedWM38 晶圆图的原始分辨率大小为 52×52 ,将整个数据集以 4 : 1 随机分配为训练集和测试集,30 412 张用于训练,7 603 张用于测试。

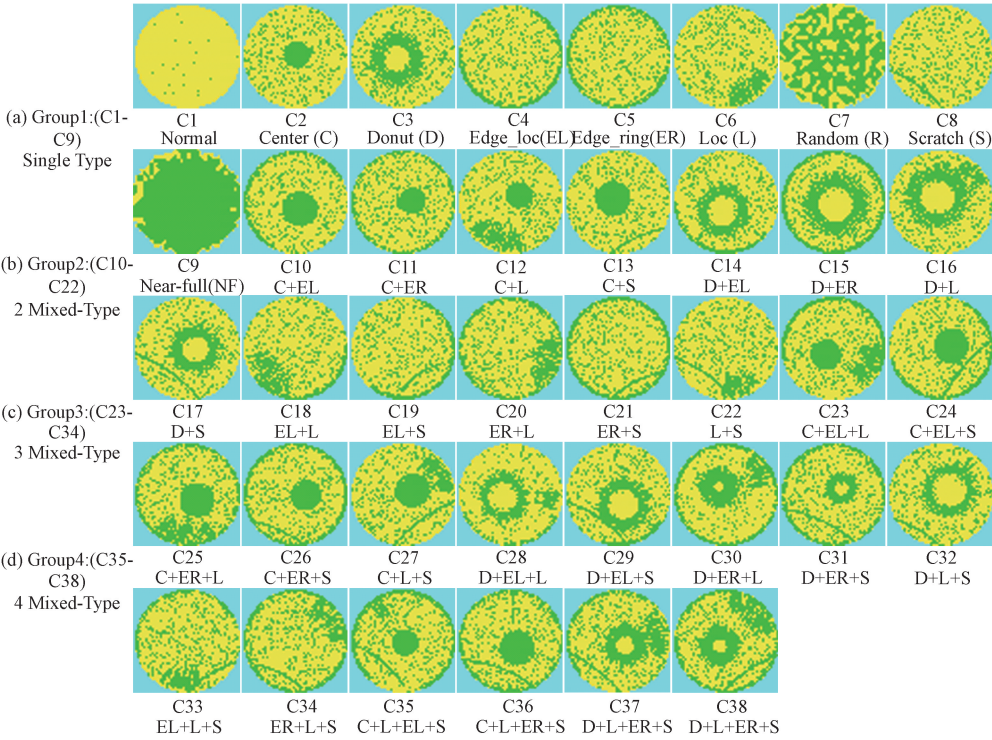


图 5 Mixed38WM 数据集集中的 38 种缺陷类型

Fig. 5 38 defect types in the Mixed38WM dataset

4 实验结果与分析

4.1 实验环境

采用 Python 环境下的 Pytorch 深度学习框架构建网络模型并完成实验。使用的 CPU 配置为 Intel (R) Core (TM) i713400F,内存为 32 GB。使用的 GPU 显卡配置为 Nvidia GeForce RTX4070,操作系统为 Windows11。所

使用的具体超参数设置见表 1。

表 1 超参数设置

Table 1 Hyperparameter settings

参数	内容
优化器	Adam
学习率	0. 001
损失函数	交叉熵损失函数
批次大小	128

为了评估模型的性能,本研究采用平均准确率(Accuracy)、精确率(Precision)、召回率(Recall)和 F1 分数等评估指标。计算平均准确率、精确率和召回率的公式见式(10)、(11)、(12)。

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

(10)

$$Precision = \frac{TP}{TP + FP} \times 100\%$$

(11)

$$Recall = \frac{TP}{TP + FN} \times 100\%$$

(12)

其中,TP、TN、FP、FN 分别代表真阳性、真阴性、假阳性、假阴性。这里的“True”和“False”表示预测是否和真实标签一致,而“Positive”和“Negative”表示根据预测是否存在相应的复合缺陷类型。F1-Score 是精确率和召回率的调和平均值,其计算公式由(13)给出。

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

(13)

4.2 对比实验

为了全面评估模型性能,与 6 种代表性先进方法进行了比较,这些方法包括 MRWA-Net、CWDR-Net、DC-Net、VGG16^[30]、MobileNetV3^[31]、DenseNet^[32],实验结果如表 2 所示。从表中可以看出,MLG-Net 在 4 个指标中都取得了最佳值,优于其他模型。这表明所提出的模型在处理混合晶圆缺陷分类任务中具有优势。

表 2 MLG-Net 与其他模型比较

Table 2 MLG-Net compared to other models %

Model	Accuracy	Precision	Recall	F1-Score
DC-Net(2020)	93.20	94.00	95.00	94.50
MRWA-Net(2023)	98.60	98.78	98.70	99.08
CWDR-Net(2023)	98.52	98.48	98.46	98.52
VGG16	98.63	98.53	98.41	98.46
MobileNetV3	97.44	97.41	97.17	97.26
DenseNet	98.72	98.52	98.59	98.54
MLG-Net	98.84	98.95	98.86	99.05

为了进一步研究 MLG-Net,本研究对比了四种晶圆缺陷分类方法(MLG-Net、MRWA-Net、CWDR-Net、DC-Net)对每类缺陷的分类准确性、精确率和召回率。结果如表 3 所示。从表中可以看出,MLG-Net 的平均分类准确率达到 98.84%,高于其他 3 种方法的 98.60%、98.52%、93.20%。此外,MLG-Net 在精确率和召回率方面也优于其他模型,分别达到了 98.95%和 98.86%,而 MRWA-Net 的精确率和召回率分别为 98.78%、98.70%。CWDR-Net 的精确率和召回率为 98.48%和 98.46%。以上数据表明,MLG-Net 在各方面的性能都优于其他方法。

从表 3 中还可以看出,针对单一缺陷类型 C1-C9,

MLG-Net 在大多数情况下都取得了较好的结果,平均准确率(Accuracy)达到 99.22%,除 C4、C7 和 C9 外,其余 6 种单一缺陷的准确率达到 100%。平均精确率(Precision)达到 98.44%,其中 C1、C7 的精确率达到 100%。而表现次优的 MRWA-Net 的平均准确率为 99.22%,与 MLG-Net 方法持平,但其平均精确率为 98.22%,比所提出模型低 0.22%。对于召回率(Recall),4 个模型都达到了 95%以上。总体而言,针对单一类型缺陷,MLG-Net 方法性能最好,MRWA-Net 次之,接下来是 CWDR-Net 和 DC-Net。

随着缺陷种类的增多,对于 2 种、3 种及 4 种混合型缺陷,MLG-Net 依旧有着很高的精确率。在 C37(D+L+EL+S)4 种缺陷上表现优异,准确率、精确率和召回率分别为:99%、98%、99%。表明即使面对更加复杂的缺陷组合,MLG-Net 仍能较好的识别缺陷并且进行准确的分类。

从实际技术意义来看,Accuracy、Precision 和 Recall 的提升很有必要,这种精细化的改进对于高度精密的晶圆批量缺陷检测尤为重要。特别在复杂的晶圆制造过程中,缺陷的种类繁多且形态多样,这要求检测算法能够在各种情况下保持较高的性能。目前最难识别的是 4 种四重混合缺陷(图 5 中的 C35-C38)。本研究对比了不同模型识别 4 种四重混合缺陷时的准确性、精确率和召回率。实验结果如表 4 所示。从表中可以看出,该模型表现依然是最好的,准确性、精确率和召回率分别达到了 97.18%、98.95%、97.16%。这表明相比现有先进方法,所提模型在面对 4 种四重混合缺陷时,具有更强的泛化能力和适应性,能够在更广泛的条件下保持较高的检测准确性。这种性能提升意味着网络能够更好地捕捉到微小或边缘的缺陷,减少漏检或误检的概率,从而确保产品的质量稳定性。

综合上述实验结果和分析,可以证明本研究提出的 MLG-Net 针对多种缺陷类型的分类性能都优于现有代表性先进方法。

4.3 消融实验

MLG-Net 模型为多分支架构,本研究开展了消融实验来证明各个分支的有效性。将全局分支(GB)作为 Baseline,将多尺度通道注意力模块(MSCA)及特征增强自注意力模块(FESA)作为两个模块。实验结果如表 5 所示。

从表 5 可以看出,当同时使用两个模块时,模型各项指标都达到了最佳值,说明 MSCA 和 FESA 模块相互配合,能够为模型提供更强大的特征提取和分类能力,从而提升了 MLG-Net 的整体性能。当单独使用 MSCA 或 FESA 模块时,模型的性能优于 Baseline,但结果低于同时使用两个模块的模型。这表明,MSCA 和 FESA 模块各自对模型性能都有所提升,但是提升性能不如同时使用两

表 3 MLG-Net、MRWA-Net、CWDR-Net 和 DC-Net 的实验结果

Table 3 Experimental results of MLG-Net、MRWA-Net、CWDR-Net and DC-Net

Model	MLG-Net			MRWA-Net			CWDR-Net			DC-Net		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Defect	C1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.94	0.91
	C2	1.00	0.97	1.00	1.00	0.99	1.00	0.98	0.99	0.98	0.93	0.97
	C3	1.00	0.98	0.99	0.99	0.97	1.00	0.97	0.98	0.97	0.95	0.93
	C4	0.97	0.99	0.97	1.00	0.96	1.00	0.97	0.98	0.97	0.94	0.91
	C5	1.00	0.98	1.00	0.97	0.98	0.97	0.97	1.00	0.97	1.00	0.97
	C6	1.00	0.98	1.00	0.99	0.99	0.99	0.96	1.00	0.96	0.94	0.99
	C7	0.99	1.00	0.99	0.99	1.00	0.99	0.89	0.92	0.89	0.96	0.94
	C8	1.00	0.98	1.00	1.00	0.98	1.00	0.98	0.98	0.98	0.93	0.88
	C9	0.97	0.98	0.99	0.99	0.97	1.00	0.99	0.98	0.99	1.00	0.97
	C10	0.98	0.99	0.99	0.98	0.97	0.98	0.96	0.97	0.96	0.99	0.94
	C11	1.00	0.98	0.99	0.99	0.99	0.99	0.96	0.99	0.96	0.98	0.92
	C12	1.00	1.00	1.00	0.99	0.99	0.99	0.98	0.99	0.98	0.99	0.92
	C13	0.99	1.00	0.99	0.99	0.98	0.99	0.98	0.98	0.98	0.97	0.89
	C14	0.99	0.98	0.99	0.99	1.00	0.99	0.95	0.99	0.95	0.99	0.96
	C15	0.99	0.99	0.99	1.00	0.99	1.00	0.97	1.00	0.97	0.96	0.91
	C16	1.00	0.99	1.00	0.95	0.98	0.95	0.89	0.97	0.89	0.98	0.94
	C17	1.00	0.97	1.00	1.00	0.98	1.00	0.98	0.95	0.98	0.93	0.96
	C18	0.99	0.98	1.00	0.98	1.00	0.98	0.94	0.98	0.94	0.94	0.89
	C19	0.98	0.98	0.97	0.97	0.97	0.97	0.98	0.98	0.98	0.92	0.94
	C20	0.98	1.00	0.98	0.97	0.98	0.97	0.96	0.98	0.96	0.95	0.91
	C21	1.00	0.98	1.00	0.99	0.96	0.99	1.00	0.97	1.00	0.91	0.96
	C22	0.97	0.98	0.97	0.97	0.98	0.97	0.99	0.93	0.99	0.90	0.88
	C23	0.97	0.98	0.96	0.95	0.99	0.95	0.97	0.96	0.97	0.89	0.99
	C24	0.99	1.00	1.00	1.00	0.99	1.00	1.00	0.99	1.00	0.89	0.92
	C25	0.99	0.97	0.99	0.99	0.97	0.99	0.97	0.97	0.97	0.91	0.93
	C26	0.99	0.97	1.00	0.99	0.96	0.99	0.98	0.98	0.98	0.93	0.97
	C27	0.96	0.99	0.96	0.97	0.99	0.97	0.98	0.95	0.98	0.91	0.97
	C28	0.98	1.00	0.98	0.99	1.00	0.99	0.99	0.99	0.99	0.88	0.95
	C29	0.97	0.99	0.98	1.00	0.98	1.00	0.97	0.95	0.97	0.91	0.98
	C30	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	0.99	0.92	0.89
	C31	0.99	0.96	0.99	1.00	0.99	1.00	0.98	0.96	0.98	0.92	0.90
	C32	0.97	1.00	0.95	0.96	0.98	0.96	0.95	0.89	0.95	0.88	0.99
	C33	0.96	1.00	0.95	0.96	0.98	0.96	0.98	0.94	0.98	0.86	0.97
	C34	0.99	0.98	0.99	0.98	0.98	0.98	0.97	0.96	0.97	0.89	0.98
	C35	0.95	1.00	0.94	0.95	0.99	0.95	0.96	0.97	0.97	0.87	0.96
	C36	0.97	0.97	0.97	0.95	1.00	0.95	0.97	0.95	0.97	0.91	0.99
	C37	0.99	0.98	0.99	0.98	0.98	0.98	0.98	0.97	0.98	0.86	0.95
	C38	0.97	1.00	0.99	0.98	0.99	0.98	0.98	0.96	0.98	0.88	0.92
	Average/%	98.84	98.95	98.86	98.60	98.78	98.70	98.52	98.48	98.46	93.20	94.00

表 4 针对 4 种四重混合缺陷不同模型的性能比较

Table 4 Performance comparison of different models for four quadruple mixed defects				%
Model	Accuracy	Precision	Recall	
DC-Net (2020)	88.13	95.60	94.10	
CWDR-Net (2023)	97.12	96.21	97.16	
MRWA-Net (2023)	96.43	98.68	96.49	
MLG-Net	97.18	98.95	97.16	

表 5 消融实验

Table 5 Ablation experiments							%
Baseline	MSCA	FESA	Accuracy	Precision	Recall	F1-Score	
✓	×	×	97.98	97.66	98.01	97.81	
✓	✓	×	98.21	98.19	98.22	98.14	
✓	×	✓	98.32	98.37	98.44	98.39	
✓	✓	✓	98.84	98.95	98.86	99.05	

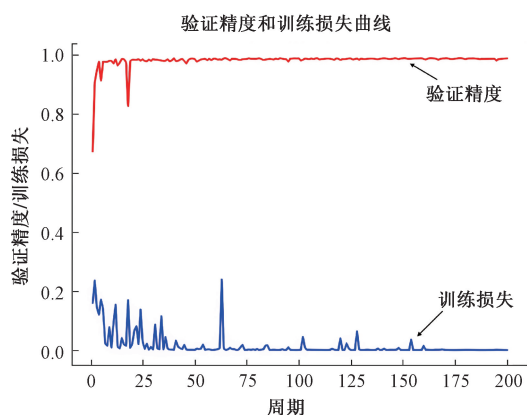


图 7 模型训练损失和验证准确性曲线

Fig. 7 Training loss and validation accuracy curves for the model

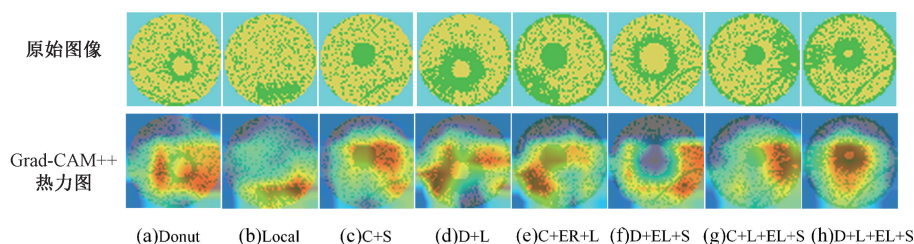


图 8 模型对 MixedWM38 数据集中八种不同晶圆缺陷类型的注意力图

Fig. 8 The model pays attention to eight different wafer defect types in the MixedWM38 dataset

缺陷交叉重叠分类问题,提出了一种名为 MLG-Net 的缺陷分类网络,通过融合全局分支与局部分支,分别提取全局语义信息、局部形状信息及局部细节信息,提高了混合型晶圆缺陷的分类精度。大量实验结果证明,MLG-Net 的性能优于现有六种代表性先进混合类型晶圆缺陷分类方法。未来的研究将致力于进一步优化网络结构,提升模型的实时性和适应性,以应对更加多样化的缺陷类型和更加复杂的工业生产环境,推动半导体行业的技术进步与产业升级。

参考文献

- [1] MONCH L, UZSOY R, FOWLER J W. A survey of semiconductor supply chain models part I: semiconductor supply chains, strategic network design, and supply chain simulation[J]. International Journal of Production Research, 2018, 56(13): 4524-4545.
- [2] ABD AL RAHMAN M, MOUSAVI A. A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry[J]. IEEE Access, 2020, 8: 183192-183271.
- [3] 姜也,黄一凡,熊美明,等. PCBA 板载 DDR 芯片焊点缺陷检测研究[J]. 仪器仪表学报, 2023, 44(2): 129-137.
- JIANG Y, HUANG Y F, XIONG M M, et al. Research on

策过程。使用 Grad-CAM++ 来分析所提出的混合型晶圆缺陷分类模型 MLG-Net。如图 8 所示,分别从每种缺陷类别中拿出两个示例,分别为(a)“Donut”、(b)“Local”、(c)“C+S”、(d)“D+L”、(e)“C+ER+L”、(f)“D+EL+S”、(g)“C+L+EL+S”、(h)“D+L+EL+S”。然后使用 Grad-CAM++ 技术生成与这些示例对应的热力图,这些热力图显示了模型在输入图像中关注的区域。从图 8 可以看出,该模型能够准确地捕捉晶圆缺陷的位置,能够将注意力集中在缺陷区域上。这表明所提模型能够有效学习到晶圆缺陷特征,并能够对它们进行准确分类。

5 结 论

针对混合型晶圆缺陷的种类、尺寸、大小等不确定且

solder bump defect detection of DDR chip on PCBA[J]. Chinese Journal of Scientific Instrument, 2023, 44(2): 129-137.

- [4] 史浩琛,金致远,唐文婧,等. 基于深度学习的高精度晶圆缺陷检测方法研究[J]. 电子测量与仪器学报, 2022, 36(11): 79-90.
- SHI H CH, JIN ZH Y, TANG W J, et al. Research on high precision wafer defect detection based on deep learning[J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(11): 79-90.
- [5] CHEN SH H, ZHANG Y X, HOU X N, et al. Wafer map failure pattern recognition based on deep convolutional neural network[J]. Expert Systems with Applications, 2022, 209: 118254.
- [6] MANIVANNAN S. Semi-supervised imbalanced classification of wafer bin map defects using a Dual-Head CNN[J]. Expert Systems with Applications, 2024, 238: 122301.
- [7] CHEN SH H, LIU M Q, HOU X N, et al. Wafer map defect pattern detection method based on improved attention mechanism[J]. Expert Systems with Applications, 2023, 230: 120544.
- [8] YU J B, SHEN Z L, WANG SH L. Wafer map defect recognition based on deep transfer learning-based densely

- connected convolutional network and deep forest [J]. Engineering Applications of Artificial Intelligence, 2021, 105: 104387.
- [9] TELLO G, AL-JARRAH O Y, YOO P D, et al. Deep-structured machine learning model for the recognition of mixed-defect patterns in semiconductor fabrication processes [J]. IEEE Transactions on Semiconductor Manufacturing, 2018, 31(2): 315-322.
- [10] WANG J L, XU CH Q, YANG ZH L, et al. Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition [J]. IEEE Transactions on Semiconductor Manufacturing, 2020, 33(4): 587-596.
- [11] NAG S, MAKWANA D, MITTAL S, et al. WaferSegClassNet-A light-weight network for classification and segmentation of semiconductor wafer defects [J]. Computers in Industry, 2022, 142: 103720.
- [12] WEI Y, WANG H. Wavelet integrated attention network with multi-resolution frequency learning for mixed-type wafer defect recognition[J]. Engineering Applications of Artificial Intelligence, 2023, 121: 105975.
- [13] LUO W, WANG H. Composite wafer defect recognition framework based on multiview dynamic feature enhancement with class-specific classifier [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12.
- [14] KANG H, KANG S. A stacking ensemble classifier with handcrafted and convolutional features for wafer map pattern classification[J]. Computers in Industry, 2021, 129: 103450.
- [15] XU Q, YU N, ESSAF F. Improved wafer map inspection using attention mechanism and cosine normalization[J]. Machines, 2022, 10(2): 146.
- [16] WANG Y, WEI Y, WANG H. A class imbalanced wafer defect classification framework based on variational autoencoder generative adversarial network [J]. Measurement Science and Technology, 2022, 34(2): 024008.
- [17] TSAI T H, WANG C Y. Wafer map defect classification using deep learning framework with data augmentation on imbalance datasets[J]. 2022. DOI:10.21203/rs.3.rs-2078809/v1.
- [18] DOSS R, RAMAKRISHNAN J, KAVITHA S, et al. Classification of silicon (SI) wafer material defects in semiconductor choosers using a deep learning ShuffleNet-V2-CNN Model[J]. Advances in Materials Science and Engineering, 2022, 2022: 1-12.
- [19] CHEN X Y, CHEN J Y, HAN X G, et al. A light-weighted CNN model for wafer structural defect detection[J]. IEEE Access, 2020, 8: 24006-24018.
- [20] SAQLAIN M, ABBAS Q, LEE J Y. A deep convolutional neural network for wafer defect identification on an imbalanced dataset in semiconductor manufacturing processes [J]. IEEE Transactions on Semiconductor Manufacturing, 2020, 33(3): 436-444.
- [21] WANG F K, CHOU J H, AMOGNE Z E. A deep convolutional neural network with residual blocks for wafer map defect pattern recognition [J]. Quality and Reliability Engineering International, 2022, 38(1): 343-357.
- [22] YOON S, KANG S. Semi-automatic wafer map pattern classification with convolutional neural networks [J]. Computers & Industrial Engineering, 2022, 166: 107977.
- [23] CHEN SH H, ZHANG Y X, YI M L, et al. AI classification of wafer map defect patterns by using dual-channel convolutional neural network [J]. Engineering Failure Analysis, 2021, 130: 105756.
- [24] CHA J, JEONG J. Improved U-Net with residual attention block for mixed-defect wafer maps[J]. Applied Sciences, 2022, 12(4): 2209.
- [25] WEI Y, WANG H. Mixed-type wafer defect recognition with multi-scale information fusion transformer[J]. IEEE Transactions on Semiconductor Manufacturing, 2022, 35(2): 341-352.
- [26] SHIM J, KANG S. Learning from single-defect wafer maps to classify mixed-defect wafer maps [J]. Expert Systems with Applications, 2023, 233: 120923.
- [27] WANG J L, GAO P J, ZHANG J, et al. Knowledge augmented broad learning system for computer vision based mixed-type defect detection in semiconductor manufacturing [J]. Robotics and Computer-Integrated Manufacturing, 2023, 81: 102513.
- [28] BYUN Y, BAEK J G. Image synthesis with single-type patterns for mixed-type pattern recognition on wafer bin maps [C]. 2022 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC). IEEE, 2022: 39-43.
- [29] 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.
- [30] LIU B, ZHANG X Y, GAO ZH Y, et al. Weld defect images classification with vgg16-based neural network [C]. International forum on digital TV and wireless multimedia communications. Singapore: Springer Singapore, 2017: 215-223.

[31] HOWARD A, SANDLER M, CHU G, et al. Searching for mobilenetv3 [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1314-1324.

[32] HUANG G, LIU ZH, VAN DER MAATEN L, et al. Densely connected convolutional networks [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4700-4708.

[33] CHATTOPADHAY A, SARKAR A, HOWLADER P, et al. Grad-cam + +: Generalized gradient-based visual explanations for deep convolutional networks [C]. 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), IEEE, 2018: 839-847.

作者简介



陈晓雷 (通信作者), 2003 年于兰州大学获得学士学位, 2006 年于兰州大学获得硕士学位, 2014 年于兰州理工大学获得博士学位, 现为兰州理工大学副教授, 主要研究方向为人工智能、计算机视觉和虚拟现实。
E-mail: chenxl703@lut.edu.cn

Chen Xiaolei (Corresponding author) received his B. Sc. degree from Lanzhou University in 2003, M. Sc. Degree from Lanzhou University in 2006, and Ph. D. degree from Lanzhou University of Technology in 2014, respectively. Now he is an associate professor at Lanzhou University of Technology. His main research interests include artificial intelligence, computer vision, and virtual reality.



李正成, 2019 年于陇东学院获得学士学位, 现为兰州理工大学硕士研究生, 主要研究方向为计算机视觉。
E-mail: 1749492931@qq.com

Li Zhengcheng received his B. Sc. degree from Longdong University in 2019. Now

he is a M. Sc. candidate in Lanzhou University of Technology. His main research interest includes computer vision.



杨富龙, 2006 年于兰州大学获得学士学位, 2009 年于兰州大学获得硕士学位, 2019 年于兰州大学获得博士学位, 现为兰州理工大学副教授, 主要研究方向为超材料、智能信息处理和电磁传感。
E-mail: yangfulong1982@126.com

Yang Fulong received his B. Sc. degree from Lanzhou University in 2006, M. Sc. Degree from Lanzhou University in 2009, and Ph. D. degree from Lanzhou University in 2019, respectively. Now he is an associate professor at Lanzhou University of Technology. His main research interests include metamaterials, intelligent information processing, and electromagnetic sensing.



温润玉, 2016 年于延安大学获得学士学位, 现为兰州理工大学硕士研究生, 主要研究方向为计算机视觉和图像处理。
E-mail: wenrunyu@hotmail.com

Wen Runyu received his B. Sc. degree from Yanan University in 2016. Now he is a M. Sc. candidate in Lanzhou University of Technology. His main research interests include computer vision and image processing.



沈星阳, 2023 年于兰州理工大学获得学士学位, 现为兰州理工大学硕士研究生, 主要研究方向为计算机视觉。
E-mail: 1071579761@qq.com

Shen Xingyang received his B. Sc. degree from Lanzhou University of Technology in 2023. Now he is a M. Sc. candidate in Lanzhou University of Technology. His main research interest includes computer vision.