

基于阵列分布信息引导的密集目标检测算法

童浩^{1,2}, 吴静静^{1,2*}, 安聪颖^{1,2}¹江南大学机械工程学院, 江苏 无锡 214122;²江苏省食品先进制造装备技术重点实验室, 江苏 无锡 214122

摘要 针对工业场景下密集相似目标检测过程中,容易出现定位误差和伪目标等问题,提出一种基于阵列分布信息引导的密集目标检测算法。从密集目标图像中提取种子目标,根据目标阵列排布规则设计四方向搜索匹配策略,以种子目标的四邻域构建候选目标匹配区域,利用重索引算法更新目标位置索引,不断遍历实现所有目标的精确定位;针对相似目标检测困难的问题,在卷积神经网络前引入 Transformer 自注意力结构,提取样本间位置和类别的相关性特征,设计基于组图孪生卷积 Transformer 的分类网络,增强相邻目标图像结构化信息,实现密集相似目标的精确分类,最终完成稳健的目标检测任务。对大量密集目标图像数据集进行实验,结果表明,所提算法在精度上优于对比算法,检测分类准确率达到 98.71%,可以完整提取目标并进行精确分类。

关键词 目标检测; 密集目标; 阵列信息; 四方向搜索匹配; Transformer

中图分类号 TP391 **文献标志码** A

DOI: 10.3788/LOP231607

Dense Target Detection Based on Array Information Guidance

Tong Hao^{1,2}, Wu Jingjing^{1,2*}, An Congying^{1,2}¹School of Mechanical Engineering, Jiangnan University, Wuxi 214122, Jiangsu, China;²Jiangsu Key Laboratory of Advanced Food Manufacturing Equipment and Technology, Wuxi 214122, Jiangsu, China

Abstract This study proposes a dense target detection algorithm utilizing array distribution information guidance to address challenges related to positioning errors and false targets commonly occurring during the detection process of numerous similar targets in industrial settings. The methodology involves extracting seed targets from dense target images and implementing a four-direction search matching strategy based on target array layout rules. It forms candidate target matching regions from the surrounding four regions of the seed targets, thereby updating the target position index through a re-indexing algorithm and conducting continuous traversing to precisely position all targets. Additionally, to address the difficulty of detecting similar targets, a Transformer self-attention structure is introduced in front of the convolutional neural network to extract correlation features of positions and categories among samples. Subsequently, a classification network based on the twin convolutional Transformer is devised to enhance structured information within adjacent target images, enabling accurate classification of dense and similar targets and thereby accomplishing robust target detection tasks. Experiments are conducted on a large number of dense target image datasets, and the results show that the proposed algorithm outperforms the comparison algorithms in accuracy, achieving detection and classification accuracy of 98.71%. Therefore, it can effectively extract targets and conduct precise classification.

Key words target detection; dense target; array information; four-direction search matching; Transformer

1 引言

工业实际生产为了满足节约成本和制造工艺的要求,往往将制造的产品按照一定规则紧密排列,如晶圆

上的晶粒^[1]和 LED 灯珠^[2]。在视觉检测中,这一类对象被称为密集目标图像,目标间距小、外观特征近似、呈紧密阵列状分布,广泛应用于半导体制造自动检测、智能仓储和电子消费等各个领域。密集目标图像中包

收稿日期: 2023-06-27; 修回日期: 2023-08-01; 录用日期: 2023-10-13; 网络首发日期: 2023-10-23

基金项目: 国家自然科学基金项目(62072416, 61873246)

通信作者: *wjldy720@jiangnan.edu.cn

含的重要阵列空间位置信息,在相似目标检测中可起到关键性作用,因此基于阵列信息的密集目标检测算法具有重要的研究价值。密集目标检测方法,一直是人们关注的热点问题,也是目标检测、目标识别,以及遥感图像处理^[3-5]等领域需要深入探索的问题。

如何准确地对图像中小间距相似目标进行精确定位并去除干扰,是密集目标检测过程中的关键问题。许多学者针对这一关键问题开展了深入研究,取得了大量研究成果。部分学者通过获取图像密集目标排布信息,从而实现密集相似目标的精确检测。Xu等^[6]通过拟合晶圆切割线的直线特征构建晶粒阵列信息,从而进行有效的目标定位。Chang等^[7]利用水平和垂直灰度投影特征,获取晶粒的阵列排布信息。利用模板匹配直接对检测目标计数是另一种有效的技术;Chen等^[8]提出改进的金字塔搜索方法进行模板匹配,以滑动窗遍历图像快速有效地完成密集目标定位;李鹤喜等^[9]利用晶圆频域特征进行互相关匹配,从而完成密集分布的晶粒检测。上述密集目标检测算法均依赖于目标的灰度、几何和频域等特征对比度。然而在复杂场景下,对于密集目标图像仅利用灰度或几何信息难以将目标和背景有效区分,导致此类方法容易出现目标缺失或伪目标。

近几年来,随着深度学习技术在计算机视觉检测领域的广泛应用,大量学者对基于深度学习的目标检测算法进行了深入研究。其中,大部分的检测算法,如YOLO^[10]和Faster R-CNN^[11]目标检测算法对于稀疏目标图像都取得了较好的效果。但这一类针对稀疏目标的检测网络,卷积运算耗时,下采样过大容易丢失目标特征信息,不能保证高分辨率密集目标图像目标提取的完整性以及目标分类的准确性^[12-13]。为了提升对密集目标的检测效果,Yang等^[14]设计了一种面向密集小目标的检测算法,使用更精细的采样策略,以达到高精度的目的。但该检测模型建立在大量标签数据集上,工业场景下难以准备高质量的数据集,这给工业实

际应用带来了严重的挑战。以ResNet^[15]为代表的卷积神经网络(CNN),具有较强的局部特征信息提取能力,而ViT模型^[16-17]中的自注意力机制相比于CNN结构更能关注图像的全局上下文信息,有助于神经网络对阵列密集目标图像中各结构单元的理解和建模,未来可能应用于密集相似目标检测中。

现有方法在复杂的密集目标场景下存在定位精度差或分类错误,且不太适用于工业密集阵列目标等问题,因此本文提出一种基于阵列信息引导的密集目标检测算法。根据密集目标显著的阵列空间规律,确定搜索匹配方向与步距,提出四方向搜索匹配完成目标定位,并根据目标相邻位置关系推理得到行列索引。针对定位后相似目标容易产生类别判断错误的问题,提出一种基于组图孪生卷积Transformer的分类网络,提取密集目标阵列空间分布特征,形成目标位置信息显著的特征图,提升目标的分类精度。所提方法保证了密集目标检测的准确性,同时避免了大量训练样本的标注需求问题。

2 密集目标图像的空间分布特性

2.1 空间排布规律

在工业场景下,密集目标通常按照某种空间分布规律排列,间距小、密度大,形成一种阵列排布的密集目标图像。晶圆上存在大量形状特征相似或基本相同的子目标——晶粒,一张晶圆上含有少则几十、多则上万的晶粒子目标,这些子目标在排布上具有可重复的空间分布关系^[18]。因此,晶圆是一种典型的阵列密集相似目标图像,符合研究对象的要求。

晶粒目标在晶圆上的排布具有密集的特点,如图1(a)所示。将晶圆图像视为一种二维图像信号,对信号进行离散傅里叶变换分析,可以在频域上获得有效的特征信息。图1(a)经过变换后,可得到图1(b)这种形式的频谱图。图像中心区域代表灰度平缓过渡的低频信号,而边缘区域代表灰度变换剧烈的高频信号。

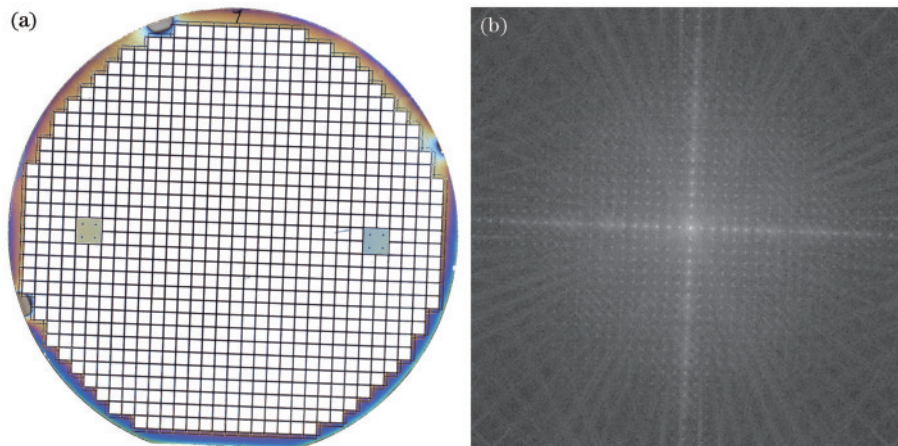


图1 密集阵列分布的晶粒目标。(a)晶圆原图;(b)频谱图

Fig. 1 Die targets with dense array distribution. (a) Original wafer image; (b) frequency spectrogram

离散傅里叶变换具有旋转性质,原图的旋转变换会呈现在频谱图上,频谱图也会旋转相同的角度。图 1(b)中灰度较高的部分代表该频率分量较强,从频谱中可以看出,晶粒具有明显的周期性规则排布特征。

2.2 目标类别与阵列特征相关性

工业密集目标图像中相似目标往往难以区分,为了实现可靠的检测,关键在于如何增强目标类别与其阵列位置的区分度特征,以实现有效识别。针对这一问题,提出一种基于子目标类型与空间分布相关性的分类方法,通过建立 3×3 阵列结构来表示目标数据类型,并提取目标类型与阵列空间分布相关性特征,形成目标位置信息显著的特征图,从而实现晶粒目标的有效区分。

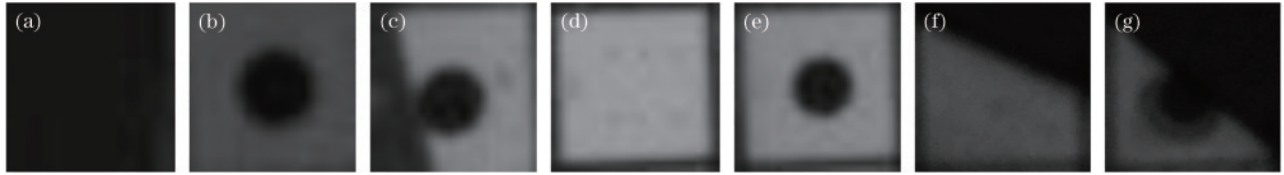


图 2 晶圆数据集各类型图片。(a)背景;(b)阴影墨点;(c)部分阴影墨点;(d)正常;(e)墨点;(f)破损;(g)破损墨点
Fig. 2 Images of various types of wafer dataset. (a) Background; (b) shaded ink dot; (c) partially shaded ink dot; (d) normal; (e) ink dot; (f) broken; (g) broken ink dot

为了更好地描述晶粒目标类别与阵列特征相关性,同时考虑到本文晶圆实验对象裕量区域宽度的限制,其阴影部分约为 3 个晶粒宽度, 3×3 阵列区域足以描述此晶圆的空间分布与类别的相关性,因此采用 3×3 结构阵列化目标数据集,如图 3 所示。晶圆上,这些目标也遵循图 3 中的位置关系,可以看出,由于裕量区域形成不同类型的晶粒,但它们在表观上十分相似,具有类间方差小的特点,仅凭借纹理特征难以有效区分。从阵列角度看,由于目标之间本身存在的相邻关

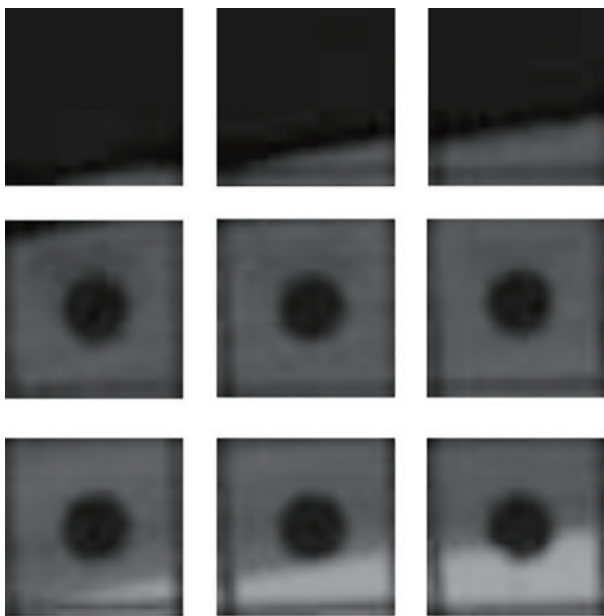


图 3 九宫格阵列目标
Fig. 3 Sudoku array target

效区分。

晶圆在制造过程为保证芯片的良率,通常会将边缘作为裕量区域^[19],通常呈现阴影特征。在芯片测试过程中,需要对不同类型的晶粒进行区分,以降低封装测试成本^[20]。由于晶圆制造工艺的影响,边缘裕量区域中的晶粒形态各异但具有相似性。本文中晶圆的晶粒类型包含背景、阴影墨点、部分阴影墨点、正常、墨点、破损和破损墨点,如图 2 所示。晶粒在裕量区域阴影覆盖程度各不相同,同类型的晶粒之间有时在灰度特征上差异较大,而不同类型的晶粒具有一定的相似性。同时,晶粒具有较少的纹理特征,因此很难准确地判定晶粒类型。

系,子目标特征会影响其相邻的目标,因此引入阵列特征信息,有助于提升晶粒分类准确率。在晶圆上,相邻的晶粒类别信息相互影响,将阵列图像组合起来,加强邻近目标图像之间的信息交互,构建阵列目标的整体结构性信息,有助于分类任务的完成。

3 密集阵列目标检测方法

提出一种基于阵列分布信息引导的密集目标检测算法,流程图如图 4 所示。所提算法包括 3 个阶段:阵列信息引导的图像预处理、基于重索引的四方向搜索匹配,以及基于组图孪生卷积 Transformer 的分类。

首先,对密集目标图像进行预处理,利用阵列信息引导图像校正以获取目标尺寸和间距(D_x, D_y)等信息,为匹配算法做准备;其次,设计基于重索引的四方向搜索匹配算法对目标进行定位和行列索引,并设计异常区域目标推理算法,利用阵列分布特征作为先验信息,对无法匹配的目标进行估计定位,为后续分类算法提供数据集支持;最后,设计组图孪生卷积 Transformer 网络(SCTAI)进行分类,使用改进的 ResNet 提取局部特征,再输入多头自注意力块中以增强阵列目标图像序列的全局特征信息,提升分类精度,从而实现目标的准确识别和检测。

3.1 阵列信息引导的图像预处理

晶圆密集目标图像的阵列信息有助于晶粒目标的检测,因此提出一种阵列信息引导的图像预处理方法。当图像中的密集目标倾斜时,目标区域匹配会产生区域误差,如图 5 所示,定位框包含背景和其他目标。为避免目标定位偏差,目标的旋转校正是必要的预处理。

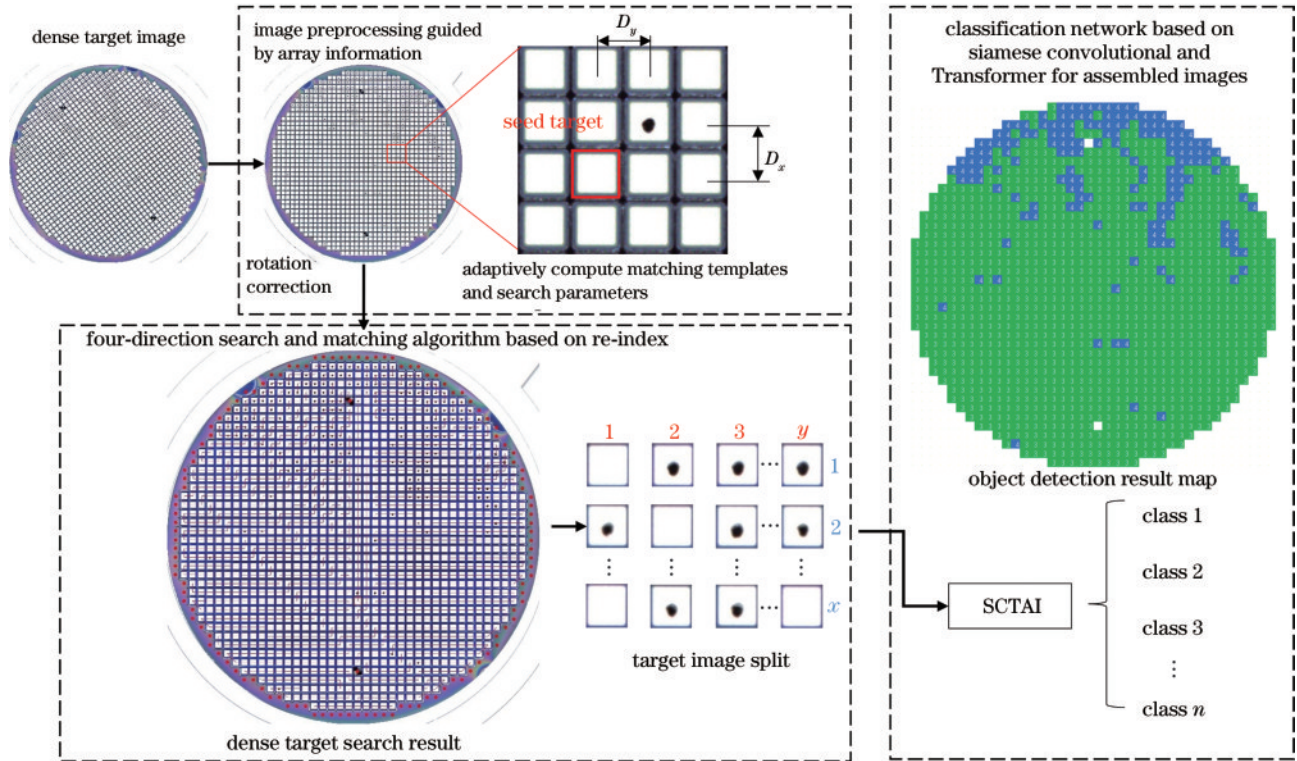


图 4 阵列信息引导的密集目标检测算法流程图

Fig. 4 Flow chart of dense target detection algorithm guided by array information

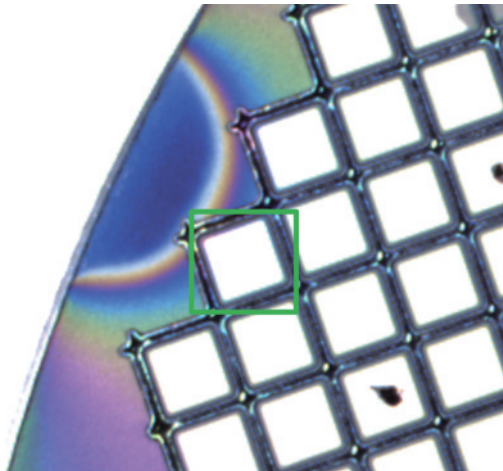


图 5 目标倾斜的定位框

Fig. 5 Positioning box of tilted target

先采用傅里叶变换对阵列目标图像[图 6(a)]进行频域变换,得到频域特征[图 6(b)]后计算图像偏转角度 α ,并对原图像进行旋转,使其恢复到正常位置,图 6(d)为角度校正后的结果。

预处理需要计算的阵列特征参数包括水平和垂直匹配步距、模板图像 A 和 B 、四方向搜索匹配区域尺寸 $H \times W$ 等。利用自适应的方法对局部晶圆图像进行处理,自动计算这些阵列结构的特征参数。选择并裁剪含有典型目标区域的图 7(a),分割得到图 7(b),再边缘修剪得到图 7(c)。进而将其划分为 m 行和 n 列,以行优先排列得到初步模板掩码 $Mask_i, i \in (1, 2, 3, \dots, mn)$ 。计算相邻行列的平均距离作为水平与垂直方向上的匹配步距,对 $Mask_i$ 进行区域扩充,保留完整的边缘,作为最终的候选模板掩码。以 $Mask_i$ 为掩模取出原图的目标图像,记为 $Blob_i$ 。在目标图像

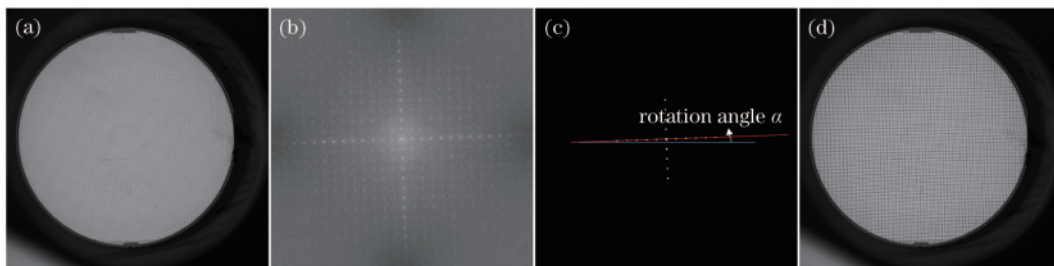


图 6 密集目标旋转校正流程。(a)原图;(b)频谱图;(c)频谱二值图角度计算;(d)旋转校正后的图像

Fig. 6 Dense target rotation correction process. (a) The original figure; (b) frequency spectrogram; (c) angle calculation of frequency spectrum binary image; (d) image after rotation correction

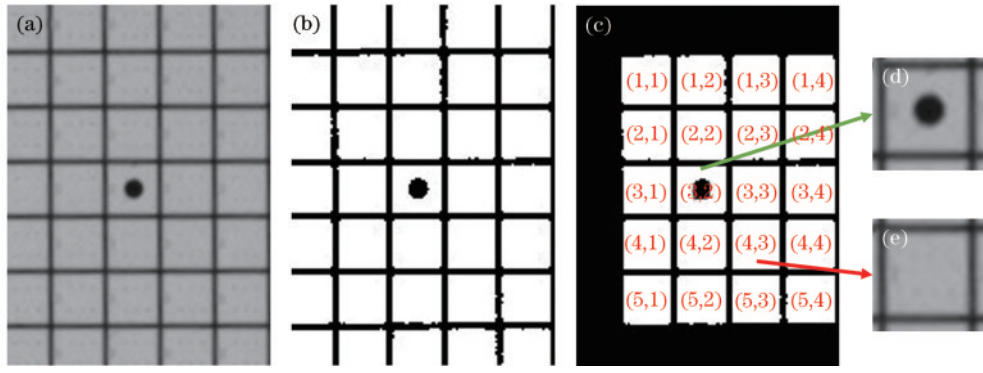


图7 自适应模板选择。(a)典型目标区域；(b)二值图；(c)裁剪后图像；(d)(e)模板A和模板B

Fig. 7 Adaptive template selection. (a) Typical target area; (b) binary image; (c) image after edge cropping; (d) (e) template A and template B

尺寸基础上增加 n 倍得到匹配区域尺寸 $H \times W$, 该倍率可以根据匹配速度和具体目标图像要求修改。对 $Blob_i$ 进行匹配, 匹配相似度 R 的计算使用标准平方差法:

$$R(T, I) = \frac{\sum_{x,y} [T(x,y) - I(x,y)]^2}{\sqrt{\sum_{x,y} T(x,y)^2 \cdot \sum_{x,y} I(x,y)^2}}, \quad (1)$$

式中: T 是模板图像; I 为待匹配图像; x 和 y 是像素坐标; R 是匹配相似度得分。将每个 $Blob$ 与其他目标的计算得到的 R 相加, 得到总相似度, $Blob_i$ 的总相似度记为 S_{simi} :

$$S_{simi} = \sum_{j=1}^{mn} [1 - R(Blob_i, Blob_j)], \quad (2)$$

式中: $i, j \in (1, 2, 3, \dots, mn)$ 。为自动获得标准的模板图像, 提高阵列密集目标提取定位精度, 当区域内主要存在两种类型的晶粒目标时, 可以选取最大和最小相似度的目标作为模板 A 和 B。

$$A = Blob_{\text{argmax}(S_{sim1}, S_{sim2}, \dots, S_{simmn})}, \quad (3)$$

$$B = Blob_{\text{argmin}(S_{sim1}, S_{sim2}, \dots, S_{simmn})}, \quad (4)$$

式中: argmax 和 argmin 函数分别输出最大值和最小值的索引。

3.2 基于重索引的四方向搜索匹配目标提取算法

在阵列分布的密集目标图像中, 目标规律均匀分布。在匹配定位过程中, 由于图像畸变、目标异常偏移和步距累计的舍入误差等问题, 采用固定步距间隔会导致定位偏差。而传统的多目标模板匹配通过滑窗遍历图像所有的位置进行定位, 其匹配次数会随着图像尺寸的增加而显著升高, 不适用于高分辨率图像。所提基于重索引的四方向搜索匹配算法可有效解决图像畸变和目标异常偏移导致的定位偏差问题。在搜索匹配的过程中, 通过将待匹配区域网格化, 提升阵列目标的定位速度, 并计算出行列位置索引。利用预处理阶段获得的阵列参数进行四方向搜索匹配, 定位中心区

域晶粒目标, 并计算出行列位置。最后针对形貌复杂的异常目标, 利用阵列分布的先验信息进行推理估计, 定位剩余目标。所提四方向搜索匹配算法具有目标索引快速和计算量小的优势, 而且对于复杂形貌目标的定位具有鲁棒性。

重索引算法^[21]利用中心种子目标的四邻域相关信息对目标进行位置索引, 邻域索引信息如图8所示。索引成功后新的目标成为中心种子目标继续索引, 不断持续搜索, 直至所有目标索引完成。具体步骤如下:

1) 选择初始的种子目标后, 搜索方向依次为上、左、右、下;

2) 根据水平和垂直匹配步距和四方向搜索匹配区域尺寸 $H \times W$, 确定4个方向的待匹配区域;

3) 判断搜索区域是否在晶圆内, 在则进入步骤4), 不在则重新进入步骤1)完成下一方向的搜索;

4) 在区域内进行模板匹配, 若相似度得分满足阈值要求, 将该目标作为新的种子目标重新进行搜索循环, 执行步骤1), 当所有目标重索引完成后, 退出循环。四方向搜索匹配的示意如图9所示, 图中心区域

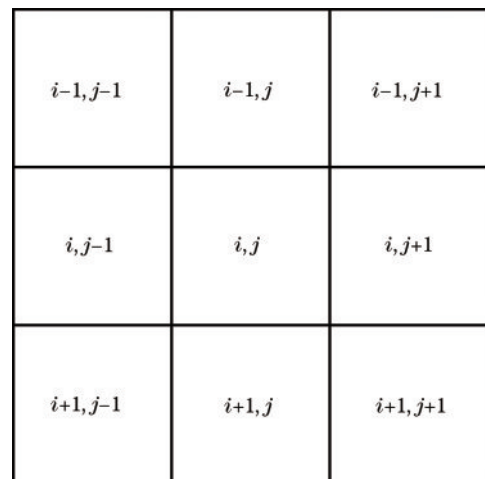


图8 九宫格重索引

Fig. 8 Sudoku re-index

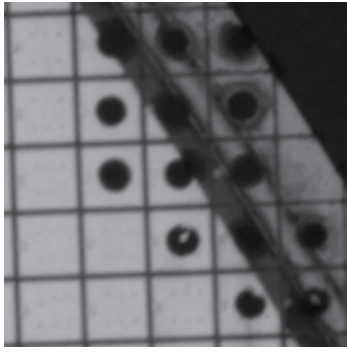


图 9 裕量区域
Fig. 9 Margin area

的无填充方框为匹配成功的目标。

在内部晶粒目标提取完成后,图 9 所示阴影部分的裕量区域未能得到很好的定位。裕量区域目标表观与模板差异较大,导致定位失效。针对上述问题,提出基于阵列引导的目标推理算法,利用密集目标的阵列分布特性进行引导,推理估计裕量区域异常目标。裕量区域的推理估计示意如图 10 所示,填充方框为估计的目标。通过搜索匹配的最外层目标与待检测区域的位置距离关系估计剩余裕量区域目标的行列数,将已匹配的目标坐标作为先验信息,推理出目标位置。

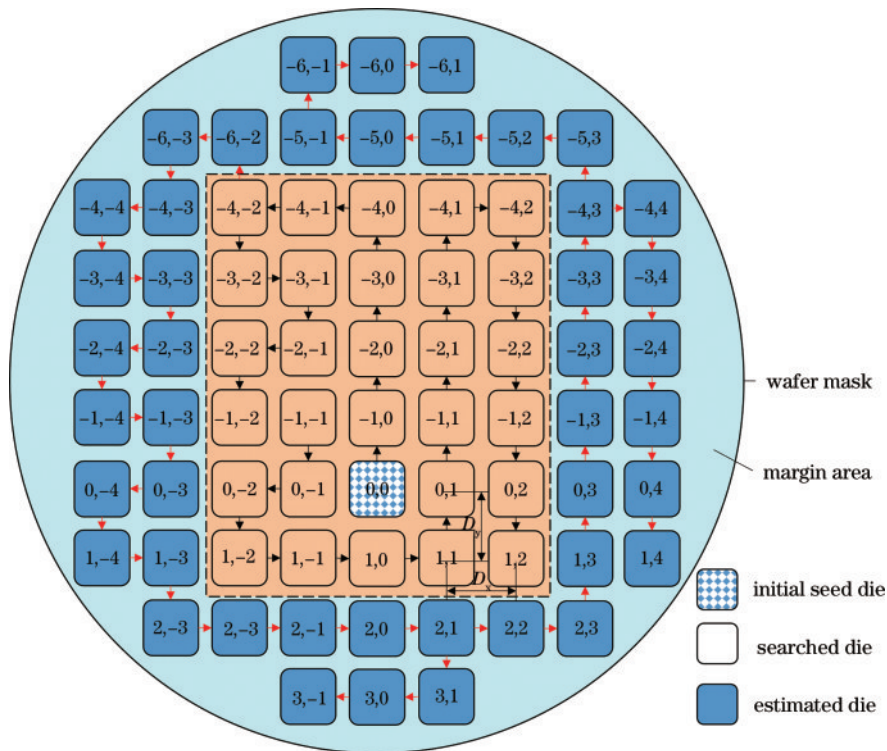


图 10 四方向搜索匹配
Fig. 10 Four-direction search matching

3.3 基于组图孪生卷积 Transformer 的分类网络

3.3.1 主干网络结构

组图孪生卷积 Transformer 主要由共享权重的孪生 ResNet(SR)特征提取层和组图 Transformer 增强模

块(TAI)组成,整体网络结构如图 11 所示。

SCTAI 主要包含 3 层,通过孪生卷积模块完成局部特征提取,进而嵌入多张图像的位置和类别信息,通过 Transformer 增强后完成分类。SCTAI 的输入与一般的

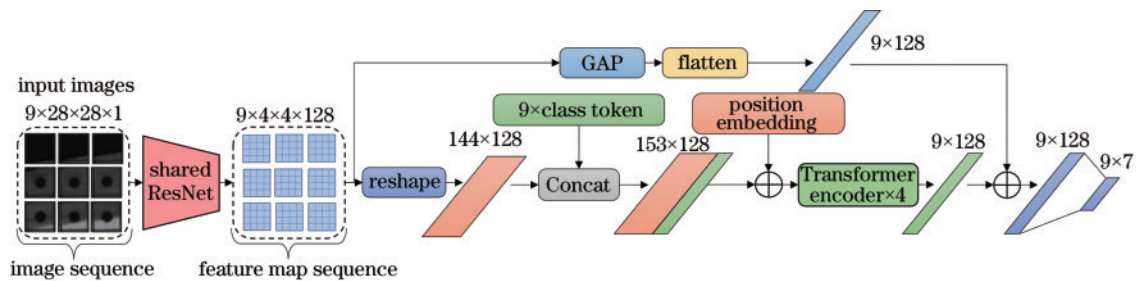


图 11 SCTAI 整体网络结构
Fig. 11 SCTAI overall network structure

分类网络的单张输入不同,是由 3×3 阵列目标组成的序列,共 9 张图像。第 1 层为了处理组合图像序列这一特殊输入类型,因此首先设计 SR 作为特征提取器,提取各目标局部特征。输入图像的序列长度为 9,输入尺寸为 $28 \times 28 \times 1$,经过特征提取后得到 9 张大小为 $4 \times 4 \times 128$ 的特征图,之后进入第 2 层。第 2 层,首先对 9 张特征图进行拼接,将拼接后的特征图 reshape 成向量,尺寸为 144×128 ,再拼接图像序列长度数量的 9 张 class token 信息,得到尺寸为 153×128 的特征向量,嵌入位置信息(position embedding)后进入第 3 层。第 3 层对来自 9 张图像的特征图与其 class token 的融合特征使用 4 层 Transformer encoder 模块进行增强,增强后的特征与 SR 得到的特征图经过平均池化和展平后进行特征相加,最后通过全连接层输出序列中各图像的最终预测结果。SR 层采用改进的共享权重 ResNet 对各个图像进行局部特征提取。TAI 层在权值共享的 CNN 的基础上,结合多张图像输入的情况,嵌入不同阵列位置的图像类别与位置信息,最后引入多头自注意力增强多张高维特征图各单元相互之间的依赖关系。在网络结构中,设计残差结构,有效提高模型训练速度并防止梯度消失。SCTAI 通过在网络的浅层位置提取图像的局部特征信息,进而对多张图像的类别位置信息进行编码,最后在网络深层对阵列结构化的信息进行全局建模。多张阵列结构图像同时输入结构以及 CNN 和 Transformer 的有效结合,使得网络能够很好地捕获图像序列的局部和全局语义信息,对相似目标分辨能力更强。

3.3.2 孪生 ResNet 特征提取层

为了提取序列中各图像的局部特征信息,在 SCTAI 中引入共享权重的孪生 ResNet。孪生 ResNet 使得序列中的图像共享同一个特征提取网络,有效地减小了模型的参数量,提高模型的泛化能力。孪生网络结构一般处理两个不同输入,用于比较输入之间的相似情况^[22]。而本研究为了适用于图像序列的输入结构,设计孪生 ResNet 同时获取各个图像的局部语义信息,将图像序列经过组合后同时输入孪生 ResNet 模块

中,完成特征提取,相较于常规的单输入网络类型,孪生结构可以同时处理多张图像信息,提升了模型正则化效果,并显著降低计算量和运算成本。孪生 ResNet 层相关公式为

$$SR(image_i) = ResNet(image_i), \quad (5)$$

式中: $image_i$ 代表输入图像序列, $i \in (1, 2, 3, \dots, n)$ 。

由于晶粒图像尺寸较小、纹理特征少,为防止目标有效信息丢失,特征提取网络在 ResNet 的基础上做了以下改进:首先,将第 1 个卷积层的卷积核大小由 7 改为 3;其次,删除 layer 1 的最大池化层;最后,放弃 ResNet 的 block 3 之后的层以保证特征的空间分辨率。修改后的 ResNet 特征提取编码器的模型参数量显著降低了,同时提升了对小尺寸目标的特征提取能力,具体结构见表 1。

表 1 ResNet 特征提取网络结构
Table 1 ResNet feature extraction network structure

Layer name	Kernel size	Stride	Channels	Repeat times	Feature map size
Conv1	3×3	1	1→32	1	$28 \times 28 \times 32$
Resblock 1	3×3	2	32→32	1	$14 \times 14 \times 32$
Resblock 2	3×3	2	32→64	1	$7 \times 7 \times 64$
Resblock 3	3×3	2	64→128	1	$4 \times 4 \times 128$

3.3.3 组图 Transformer 增强模块

组图 Transformer 增强模块主要采用 Vision Transformer 的网络结构,利用其多头自注意力模块提升网络对密集阵列目标的长距离特征提取能力。由于不同于单图像输入的 Vision Transformer 结构,为了同时完成整个序列各图像的分类,将 SR 得到的 9 个特征图记为 $(f_1, f_2, f_3, f_4, f_5, f_6, f_7, f_8, f_9)$,需要增加每张图像的 class token,记为 $CT_i, i \in (1, 2, 3, \dots, 9)$ 。同时为了表征整个特征序列中相对位置信息,需要加入 position embedding。具体结构如图 12 所示。

自注意力机制通过 Q, K, V 特征向量计算图像序列特征之间的相似度,增强图像的长距离语义信息。

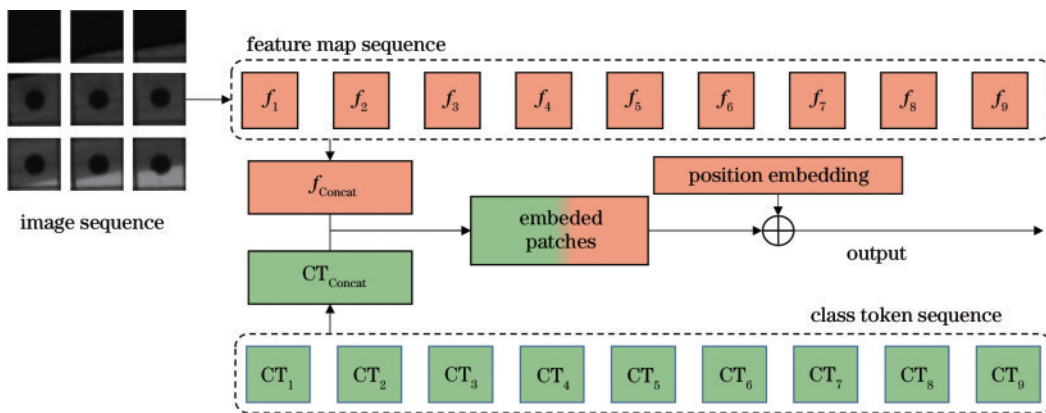


图 12 特征图序列类别与位置特征编码网络结构

Fig. 12 Feature map sequence category and location feature coding network structure

多头注意力机制在原始自注意力机制的基础上,对计算出来的注意力进行划分,使得网络可以关注多个不

同维度的特征信息,提升网络特征提取能力,其结构如图 13 所示。

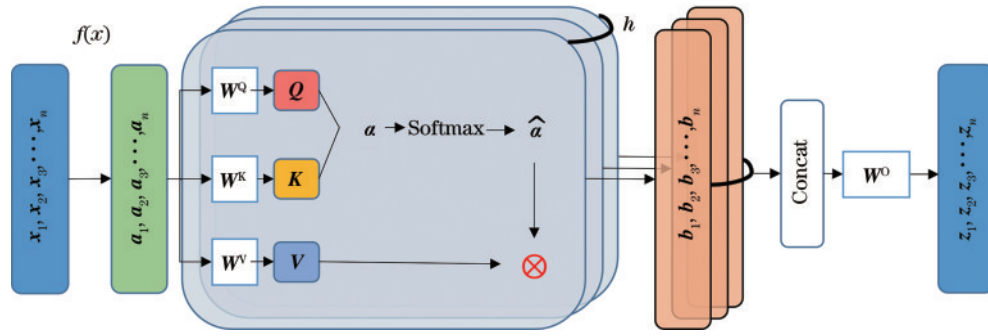


图 13 多头自注意力机制网络结构

Fig. 13 Network structure of multi-head self-attention mechanism

在图 13 中, $x_i, i \in (1, 2, 3, \dots, n)$ 为图像经过切分的 patch 序列, 通过 $f(x)$ 即线性映射到高维的 $a_i, i \in (1, 2, 3, \dots, n)$ 。将 a_i 与 W^Q, W^K, W^V 权重矩阵进行相乘得到 Q, K, V (查询矩阵、键值矩阵、值矩阵), 再通过 Q 与 K^T 相乘得到 α , 经过 Softmax 函数后得到 $\hat{\alpha}$, 再与 V 相乘得到输出自注意力向量 $b_i, i \in (1, 2, 3, \dots, n)$, 计算公式如式(6)所示。而多头自注意力机制是在 Attention 的基础上组合而成的, 将 a_i 乘以 h 个不同的权重 W^Q, W^K, W^V 可以得到 h 组 Q, K, V 。最后将经过计算后的 h 组 Q, K, V 拼接, 再乘以权重矩阵 W^O 得到 MultiHead (Q, K, V), 输出序列为 $z_i, i \in (1, 2, 3, \dots, n)$, 计算公式如式(7)所示。

$$\text{Attention}(Q, K, V) = \text{Softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (6)$$

$$\text{MultiHead}(Q, K, V) =$$

$\text{Concat}(\text{Attention}_1, \text{Attention}_2, \dots, \text{Attention}_h)W^O, \quad (7)$
式中: d_k 为键的维度。

将通过特征提取器提取后的 9 张 feature maps 展平后拼接再与 9 个 class token 进行拼接, 加上位置编码后输入多头自注意力模块中, 通过该结构每张图片的 class token 不仅受到自身局部特征信息的约束, 而且受到整个组图 token 序列上下文信息的引导, 增强后的 class token 与 feature map 的池化结果相加构建残差结构防止网络退化, 最后通过分类头全连接后得到最后的输出。

4 实验分析

4.1 实验数据

本实验使用某晶圆封装厂商提供的晶圆数据集作为实验数据。该类型晶圆图像分辨率达到 5120×5120 , 共有 20891 个晶粒, 晶圆图像如图 6(a) 所示。晶圆图像通过四方向搜索匹配方法得到共计 20905 张晶粒图像(含背景图像), 分辨率为 28×28 , 包含 7 种晶粒类别, 各类别示例图像如图 2 所示。

由于所提网络需要 9 张相邻九宫格图片作为一组

阵列图像输入, 为保证对比实验数据集完全相同以及数据集类别均衡, 需要对整个图像数据集进行随机不重合抽取九宫格图像之后重新划分数据集。划分后共生成互不重合的 346 组九宫格图像, 将 346 组图像按照 7:3 的比例进行划分, 训练集 243 组共 2187 张图像, 测试集 103 组共 927 张图像以验证网络准确率, 此为取整后的结果。

主要进行密集目标提取效果与分类准确率的实验分析, 验证四方向搜索匹配方法的效率、提取效果, 以及基于组图孪生卷积 Transformer 分类网络的可行性和高效性。目标提取实验确定晶粒模板尺寸为 30×30 , 四方向搜索感兴趣区域(ROI)尺寸为 34×34 。分类网络部分进行网络结构设计有效性分析, 对比实验选择 ResNet18、MobileNetV2、ViT 和传统方向梯度直方图(HOG)结合特征支持向量机(SVM)的分类方法。主要进行的 5 组实验包括: 1) 阵列分布密集目标提取效果对比实验分析, 对比分析四方向搜索匹配重索引算法与普通模板匹配, 以及通用目标检测 YOLOV5 算法效果; 2) 3×3 阵列输入结构与编码器改进有效性实验分析; 3) 不同分类方法在晶圆数据集上的实验结果对比; 4) 所提方法在晶粒分类数据集的参数量和计算量; 5) 晶粒图像分类误差分析。

4.2 实验环境与超参数设置

本实验计算机的 CPU 为 i9-12900K, 主频为 3.60 GHz, GPU 为 RTX3090, 显存大小为 24 GB, 内存为 32 GB。网络模型实现基于 PyTorch 框架。网络训练采用 Adam 优化器进行优化, 所有模型参数采用相同的学习率进行训练, 初始学习率为 0.01, 引入余弦退火的学习下降方式, 迭代次数设置为 200, batch_size 设置为 16, 图像分辨率设为 28×28 。

4.3 实验结果分析

4.3.1 阵列分布密集目标提取效果对比实验分析

为了验证四方向搜索匹配的重索引算法在目标提取上的效率和有效性, 将所提匹配算法与多目标模板匹配算法和 YOLOV5 进行对比, 实验结果如表 2

所示。由表 2 可知,所提算法与其他算法相比具有明显的性能提升。多目标模板匹配算法对于形貌复杂的区域无法提取,而所提算法通过阵列分布信息对异常目标区域进行推理估计,有效解决了单一模板匹配算法在复杂形貌区域无法有效提取目标的问题。此外,YOLO 对于晶圆这种高分辨率目标密集排布的图像检测效果较差,晶粒目标小且特征不明

显,降采样压缩导致原有的位置和语义信息损失,同时目标过于密集导致检测框发生重叠,检测结果不佳。多目标模板匹配算法提取的目标无相邻关系,需应用行列聚类等方法构建目标行列索引,会增加算法复杂度,而所提算法运用重索引的思路,在搜索匹配过程中提取目标四邻域信息,匹配完成后自动完成阵列的行列索引。

表 2 不同提取算法在图 6 晶圆数据集上的性能对比

Table 2 Performance comparison of different extraction algorithms on the wafer dataset of Fig.6

Matching algorithm	Miss ratio /%	Number of extracted targets	Running time /ms
Multi-target template matching	8.80	19053	1512
YOLOV5x	73.60	5515	123
Four-direction search matching algorithm based on re-index	0.00	20905	884

为直观展示各算法的提取性能,选用密集目标的局部图像进行展示,图 14 为多目标模板匹配、YOLOV5x 和所提算法的检测结果。传统的多目标模板匹配算法对于异常形貌区域存在漏检。而 YOLOV5x 算法则存在更为严重的漏检和定位误差,这是由于 YOLO 等检测网络在超高分辨率下对于小目标检测能力较弱,往往将图片进行降采样之后送入网络,这加剧了小目标语义位置信息丢失的问题。而基于重索引的四方向搜索匹配算法将模板匹配与推理估计相结合,在存在异常区域的阵列分布密集目标上取得了较好的性能。实验结果表明,所提算法利用阵列信息作为引导,较多目标模板匹配定位和通用目标检测 YOLO 算法有更高的精度,在晶圆密集目标提取效果上表现更佳。

4.3.2 3×3 阵列输入结构与编码器改进有效性实验分析

针对分类网络模型输入结构及特征提取编码器设计的有效性,将 1×1 结构、3×3 结构与 5×5 结构进行对比实验,结果见表 3。从表中可以看出:1×1 阵列结构,即单张图片输入,其分类准确率最低,无法表征阵列结构信息,仅凭一张图像不能够达到很好的分类效果;而 5×5 阵列结构相比于 3×3 阵列结构准确率下降 0.52 百分点。同时原始 ResNet18 和 ResNet34 结合 3×3 阵列结构的精度优于 5×5 阵列结构,其中,ResNet18 的 5×5 结构相比于 3×3 结构在分类准确率上下降 0.91 百分点。3×3 阵列的分类准确率最佳,可达到 98.71%,可能是因为晶圆对象裕量区域宽度约为 3 个晶粒距离,可更好地表达此晶圆的阵列结构与类型相关性,过大的阵列尺寸会导致信息的干扰,影响分类精度。综合以上考虑,选择 3×3 阵列作为输入结构。本文在孪生编码器部分也进行了改进,在 ResNet18 的基础上,针对小尺寸的晶粒目标,对原网络进行了结构优化,分别采用改进的 ResNet 与 ResNet18 和 ResNet34 作为孪生编码器进行对比实验。从表 3 可以看出,ResNet18+TAI 和 ResNet34+TAI 结构在分类精度上和 F1 分数均不如改进后的 ResNet+TAI。改进后的 ResNet 使用较浅的特征提取层,删除最大池化,保证了目标细节信息的完整性,性能最佳。

表 3 不同网络结构在晶圆数据集上的性能对比

Table 3 Performance comparison of different network structures on wafer dataset

Structure	Accuracy /%	F1 score
1×1(proposed ResNet+TAI)	98.06	0.9652
3×3(proposed ResNet+TAI)	98.71	0.9746
5×5(proposed ResNet+TAI)	98.19	0.9617
3×3(ResNet18+TAI)	98.59	0.9680
3×3(ResNet34+TAI)	97.73	0.9506
5×5(ResNet18+TAI)	97.68	0.9556
5×5(ResNet34+TAI)	97.68	0.9474

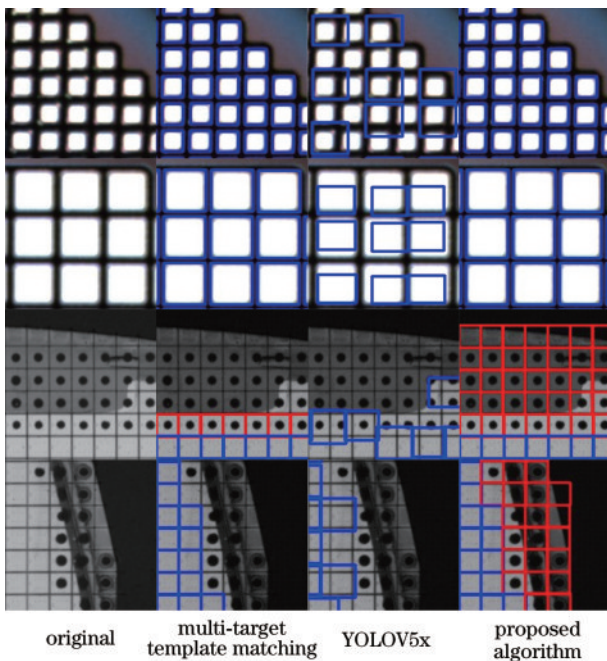


图 14 多目标模板匹配、YOLOV5x 和基于重索引的四方向搜索匹配算法的主观结果对比

Fig. 14 Comparison of subjective results of multi-target template matching, YOLOV5x, and four-direction search matching algorithm based on re-index

4.3.3 不同分类方法在晶圆数据集上的实验结果对比

为了验证基于组图孪生卷积 Transformer 的分类网络在晶粒数据集上分类的准确率与有效性,分别将 SCTAI 方法与深度学习网络 ResNet18、MobileNetV2、ViT 和 HOG+SVM 进行对比实验,实验结果如表 4 所示。从表 4 可知,所提方法与其他分类模型相比准确率取得较明显的提升,在晶圆晶粒数据集上达到了最高的分类准确率,为 98.71%。在其他方法中,ResNet18 分类网络达到了最高分类准确率,为 98.17%,所提方法相比其提升了 0.54 个百分点。ViT 模型在晶圆数据集上分类效果不佳,因为 ViT 模型参数过多,适合 ImageNet 等大型数据集任务,而对晶粒的分类准确率较低。HOG+SVM 分类器经过超参数选优后,使用高斯核函数,惩罚因子为 10, gamma 参数为 0.001,此时取得最高准确率 92.88%,传统特征提取方法对于这种相似目标特征表达能力不足,手工 HOG 特征难以表征复杂的语义特征,导致分类效果差。MobileNetV2 和 ResNet18 准确率都低于所提方法,是由于针对近似目标传统 CNN 分类网络通过卷积只能获取局部特征信息,同时对纹理特征少的小目标降采样容易丢失关键的语义信息,因此分类效果较差。所提方法先使用共享的 CNN 特征提取器,有效降低整体网络参数,再通过多头自注意力模块对提取的局部特征进行增强,建立晶粒内部以及相邻晶粒目标的长距离依赖关系,注入了相邻晶粒位置和类别相关性的先验信息。实验结果表明,所提方法相较于其他图像分类方法取得较高的准确率,更适用阵列分布下的密集目标分类场景。

表 4 不同分类方法在晶圆数据集上的性能对比

Table 4 Performance comparison of different classification methods on wafer dataset

Method	Accuracy / %	FLOPS / 10^9	Params / 10^6
ResNet18	98.17	0.2986	11.1738
MobileNetV2	98.17	0.0546	2.2322
ViT	95.58	150.7000	85.0652
HOG+SVM	92.88		
SCTAI	98.71	0.2137	1.1175

4.3.4 晶粒图像分类的参数量与计算量

图 15 展示了 SCTAI 分类方法训练时的损失值和准确率随迭代次数的变化曲线。在迭代次数达到 100 时,网络逐渐开始收敛,损失最终稳定在 0.005 左右。SCTAI 与其他方法在晶粒数据集上的计算量与参数量对比如表 4 所示。从表中可见,SCTAI 的计算量(FLOPS)约为 0.2137×10^9 ,相比 ViT 网络显著降低,且计算量比 ResNet18 减少约 30%。SCTAI 的参数量(Params)仅为 1.1175×10^6 ,与其他分类方法中的最少参数量相比降低了 50%,参数量明显减小。因此 SCTAI 相比于其他分类方法,有效降低了计算量与参数量,有利于模型的工业部署和满足实时检测要求。

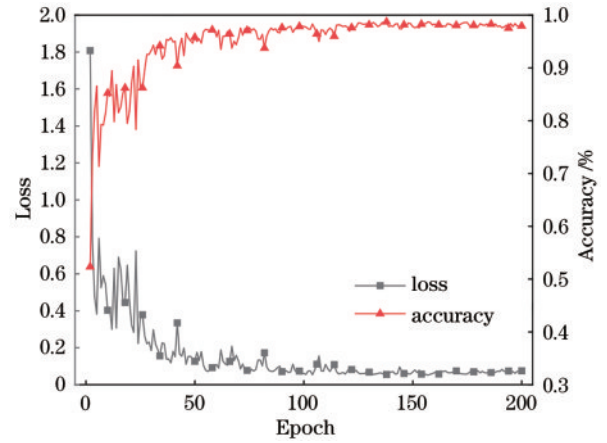


图 15 损失值和准确率训练变化曲线

Fig. 15 Loss value and accuracy training change curve

4.3.5 晶粒图像分类的误差分析

通过混淆矩阵进行晶粒分类误差分析,可以直观反映各个类别的分类情况。图 16 为所提分类网络在晶粒数据集上生成的混淆矩阵。晶粒测试集测试结果表明,绝大多数图像都集中在对角线上,表明其被正确地分类到所属类别中。所提方法对晶圆测试集的整体误差率仅为 1.29%,对正常晶粒、背景晶粒都达到了 100% 的准确率,在阴影墨点、部分阴影、墨点、破损和破损墨点的误差率分别为 5.71%、3.39%、1.12%、4.55% 和 2.78%。因此,所提分类方法在密集目标晶粒数据上达到了比较好的分类精度。

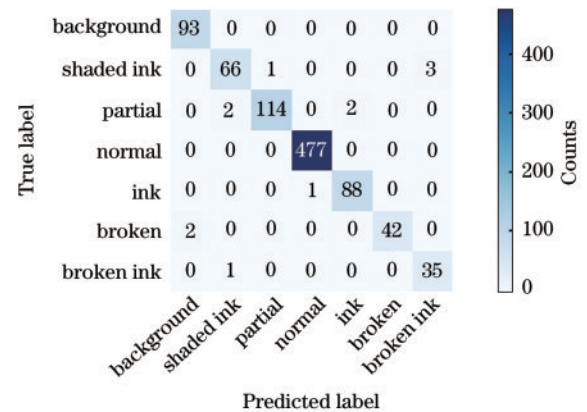


图 16 晶圆测试集上的混淆矩阵

Fig. 16 Confusion matrix on wafer test set

5 结 论

提出一种基于阵列分布信息引导的密集目标检测算法。实验结果表明,所提先定位再识别的算法可以有效完成晶圆密集目标对象检测。基于重索引的四方向搜索匹配算法相比于其他传统匹配算法和端到端的目标检测算法具有更高的定位精度,保证了晶粒目标的零漏检。基于组图孪生卷积 Transformer 的分类网络,与其他方法相比,在定位精度和速度上具有优势,获

得了 98.71% 的最高分类准确率,同时网络的参数量和计算量也有效降低,网络轻量高效。所提方法可以解决阵列分布的密集晶粒目标的检测难题,未来可以拓展到类似场景下其他密集阵列目标的检测任务中。

参 考 文 献

- [1] Yang J, Xu Y, Rong H J, et al. A method for wafer defect detection using spatial feature points guided affine iterative closest point algorithm[J]. *IEEE Access*, 2020, 8: 79056-79068.
- [2] Wang Z Y, Gong S H, Li D L, et al. Error analysis and improved calibration algorithm for LED chip localization system based on visual feedback[J]. *The International Journal of Advanced Manufacturing Technology*, 2017, 92(9): 3197-3206.
- [3] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 42(2): 318-327.
- [4] Shen Y Y, Liu D, Zhang F Z, et al. Fast and accurate multi-class geospatial object detection with large-size remote sensing imagery using CNN and Truncated NMS [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2022, 191: 235-249.
- [5] 宋子壮, 杨嘉伟, 张东方, 等. 基于无监督域适应的低空海面红外目标检测[J]. *光学学报*, 2022, 42(4): 0415001. Song Z Z, Yang J W, Zhang D F, et al. Low-altitude Sea surface infrared object detection based on unsupervised domain adaptation[J]. *Acta Optica Sinica*, 2022, 42(4): 0415001.
- [6] Xu S Q, Cheng Z H, Gao Y, et al. Visual wafer dies counting using geometrical characteristics[J]. *IET Image Processing*, 2014, 8(5): 280-288.
- [7] Chang H T, Pan R J. Automatic counting of packaged wafer die based on machine vision[C]//2012 International Conference on Information Security and Intelligent Control, August 14-16, 2012, Yunlin, Taiwan, China. New York: IEEE Press, 2013: 274-277.
- [8] Chen F J, Ye X Q, Yin S H, et al. Automated vision positioning system for dicing semiconductor chips using improved template matching method[J]. *The International Journal of Advanced Manufacturing Technology*, 2019, 100(9): 2669-2678.
- [9] 李鹤喜, 韩新乐, 杨铁牛. 基于频域互相关的 LED 晶粒阵列快速视觉检测[J]. *计量学报*, 2018, 39(4): 476-480. Li H X, Han X L, Yang T N. Rapid vision measurement of LED die array based on cross correlation in frequency domain[J]. *Acta Metrologica Sinica*, 2018, 39(4): 476-480.
- [10] Redmon J, Divvala S, Girshick R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 779-788.
- [11] Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [12] 赵菲, 邓英捷. 融合多异构滤波器的轻型弱小目标检测网络[J]. *光学学报*, 2023, 43(9): 0915001. Zhao F, Deng Y J. Light dim small target detection network with multi-heterogeneous filters[J]. *Acta Optica Sinica*, 2023, 43(9): 0915001.
- [13] Goldman E, Herzig R, Eisenschtat A, et al. Precise detection in densely packed scenes[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2020: 5222-5231.
- [14] Yang X, Yang J R, Yan J C, et al. SCRDet: towards more robust detection for small, cluttered and rotated objects[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Republic of Korea. New York: IEEE Press, 2020: 8231-8240.
- [15] 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), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.
- [16] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22)[2023-07-01]. <https://arxiv.org/abs/2010.11929>.
- [17] 袁媛, 陈明惠, 柯舒婷, 等. 基于集成卷积神经网络和 ViT 的眼底图像分类研究[J]. *中国激光*, 2022, 49(20): 2007205. Yuan Y, Chen M H, Ke S T, et al. Fundus image classification research based on ensemble convolutional neural network and vision transformer[J]. *Chinese Journal of Lasers*, 2022, 49(20): 2007205.
- [18] 李吉成, 张志龙, 鲁新平. 一种基于谱图理论的阵列目标识别方法[J]. *信号处理*, 2009, 25(10): 1527-1531. Li J C, Zhang Z L, Lu X P. A New Algorithm for Array Target Recognition Based on spectral graph theory[J]. *Signal Processing*, 2009, 25(10): 1527-1531.
- [19] Delahaye B, Baltzinger J L, Denis L, et al. Edge and extreme edge wafer manufacturing on 200 mm wafer: methodology, yield challenges, cost effective solutions, limitations[C]//2009 IEEE/SEMI Advanced Semiconductor Manufacturing Conference, May 10-12, 2009, Berlin, Germany. New York: IEEE Press, 2009: 100-105.
- [20] Kang S, Cho S, An D, et al. Using wafer map features to better predict die-level failures in final test[J]. *IEEE Transactions on Semiconductor Manufacturing*, 2015, 28(3): 431-437.
- [21] Wu T, Li B, Wang L W, et al. Automatic detectand match of LED dies basing on position relations betweenadjacent dies[J]. *Multimedia Tools & Applications*, 2010, 63(1): 107-127.
- [22] 陈志旺, 张忠新, 宋娟, 等. 基于目标感知特征筛选的孪生网络跟踪算法[J]. *光学学报*, 2020, 40(9): 0915003. Chen Z W, Zhang Z X, Song J, et al. Tracking algorithm for Siamese network based on target-aware feature selection[J]. *Acta Optica Sinica*, 2020, 40(9): 0915003.