

DOI: 10.3785/j.issn.1008-973X.2020.06.021

# 基于迁移学习与深度森林的晶圆图缺陷识别

沈宗礼, 余建波

(同济大学 机械与能源工程学院, 上海 201804)

**摘要:** 为了有效识别晶圆图缺陷模式并及时诊断制造过程的故障源, 提出基于迁移学习和深度森林集成的 DenseNet-GCForest 晶圆图缺陷模式识别模型. 为了解决深度学习模型训练困难和晶圆图缺陷类型数目不平衡的问题, 利用迁移学习将深度卷积神经网络 DenseNet 在 ImageNet 上预训练的网络权重参数迁移至本模型并重新设计分类层, 以减少深度网络模型的训练时间并提高模型的特征提取能力; 基于 DenseNet 网络提取的高维抽象晶圆图特征, 引入深度森林模型进行晶圆图特征缺陷模式识别. 工业案例的实验验证结果表明, 该方法的识别准确率达到 96.8%, 并提高了识别效率, 其性能优于典型的卷积神经网络以及其他常用识别方法.

**关键词:** 半导体制造; 晶圆缺陷; 迁移学习; 卷积神经网络; 深度森林

**中图分类号:** TP 391.41 **文献标志码:** A **文章编号:** 1008-973X(2020)06-1228-12

## Wafer map defect recognition based on transfer learning and deep forest

SHEN Zong-li, YU Jian-bo

(School of Mechanical Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** A wafer map pattern recognition (WMPR) model was proposed based on transfer learning and deep forest, in order to identify the defect pattern of the wafer maps and to timely diagnose the source of the fault in the manufacturing process. Transfer learning was used to migrate the network weight parameters of the deep CNN DenseNet pre-trained on ImageNet to this model, and the classification layer of the model was redesigned, in order to solve the problems of difficulties of deep learning model training and imbalance in the number of defect types in wafer maps. Thus, the training time of the model was reduced and the feature extraction ability was improved. Deep forest model was introduced to identify the wafer defect pattern, based on the abstract features of the wafer maps extracted by DenseNet. The experimental results on an industrial case demonstrated that the average recognition rate was about 96.8%. This method can improve the recognition efficiency and its performance is better than those well-known CNNs and other typical classifiers.

**Key words:** semiconductor manufacturing; wafer defect; transfer learning; CNN; deep forest

半导体制造已然成为现代工业最为重要的行业之一, 晶圆制造工艺通常涉及数百个集成电路模块, 由于制造过程的复杂性和动态性, 半导体晶圆容易出现缺陷<sup>[1-3]</sup>. 晶圆图上的典型缺陷图案通常是由特定的设备故障或工艺变化引起的. 例如, 边缘环形缺陷一般由错误蚀刻所引起, 线性

类缺陷是由于机器与晶圆片的摩擦, 中心类缺陷通常因薄膜沉积产生<sup>[3]</sup>. 由此可见, 快速识别产品质量异常, 并确定生产过程中的故障原因, 可以有效减少半导体制造过程中的返工率和废品率, 有助于企业提高生产效率并降低成本.

晶圆图缺陷识别通常分为缺陷检测和缺陷分

收稿日期: 2019-04-30.

网址: [www.zjujournals.com/eng/article/2020/1008-973X/202006021.shtml](http://www.zjujournals.com/eng/article/2020/1008-973X/202006021.shtml)

**基金项目:** 国家自然科学基金资助项目 (71777173); 上海科委“科技创新行动计划”高新技术领域资助项目 (19511106303); 中央高校基本科研业务费资助项目 (22120180068, 22120190196).

**作者简介:** 沈宗礼 (1996—), 男, 硕士生, 从事深度学习与故障诊断研究. orcid.org/0000-0002-0503-1709. E-mail: [979949752@qq.com](mailto:979949752@qq.com)  
通信联系人: 余建波, 男, 教授. orcid.org/0000-0003-3204-2486. E-mail: [jbyu@tongji.edu.cn](mailto:jbyu@tongji.edu.cn)

类2个步骤.传统缺陷检测算法通过对比待检测图像和参考图像,选取两者差异区域作为图像缺陷<sup>[4]</sup>,但精准的参考图像不易得到.传统的缺陷分类算法先在晶圆图中提取特征,再将该特征作为分类器的输入,从而进行模式识别<sup>[5]</sup>.上述方法主要存在以下2个问题:1)传统缺陷检测算法提取出的局部特征无法有效表示特定问题中的不同类型缺陷;2)传统缺陷检测分类算法的鲁棒性弱,需要根据新的问题重新设计整个识别模型.

随着机器学习技术的发展,各种识别器被广泛用于晶圆图缺陷识别并获得了较好的效果.监督识别分类方法有反向传播网络(back-propagation network, BPN)<sup>[6]</sup>、广义回归神经网络(general regression neural network, GRNN)<sup>[7]</sup>、支持向量机(SVM)<sup>[8-9]</sup>、随机化广义回归网络(randomized general regression network)<sup>[10]</sup>、K-临近算法(K-nearest-neighbor, KNN)<sup>[11]</sup>、决策树(CART4.5)<sup>[12]</sup>、高斯混合模型<sup>[13]</sup>、多特征决策树集成<sup>[14]</sup>、均值标准集成方法<sup>[15]</sup>等.基于机器学习的识别算法在晶圆缺陷识别中取得了一定成果,但晶圆图的特征维度高、噪声多,会严重影响识别器的性能.因此,该类方法依赖人类的主观特征选择,为识别器提供有效判别信息,这会导致所提取的特征具有一定的主观性和局限性.此外,该方法依赖人的主观经验去抽取不同类型晶圆图中的缺陷特征,因此如何保证特征的有效性是实际应用中的一大难题.综上所述,现代智能诊断系统迫切需要能客观且高效地从原始图像信号进行特征学习的方法,从而提高系统的工业可应用性.基于深度学习的特征提取方法是解决这一问题的有效手段之一<sup>[16]</sup>.

近年来,基于深度学习的特征提取方法在计算机视觉任务中获得了巨大成功<sup>[16]</sup>.深度学习也称为深度神经网络(deep neural network, DNN),是一种新型特征学习方法,具有多个隐藏层,采用具有多个神经层的分层结构,并逐层学习输入数据中的特征信息.这种深层结构能够学习复杂原始数据的高级抽象特征,在构造分类器或识别器时更容易提取有用的特征<sup>[17]</sup>.随着计算机性能的不断加强,深度神经网络实现了高效处理大规模数据并展现出了卓越的高级表征能力,因此被广泛应用在实际工业生产活动中,例如:深度置信网络(deep belief networks, DBN)、自编码器(auto encoder, AE)、降噪自编码器(denoising autoencoder, DAE)、堆叠降噪自编码器(stacked denois-

ing autoencoder, SDAE)在机器健康监测和故障诊断领域的应用引起了学者们的广泛关注<sup>[18-19]</sup>.虽然上述方法能够有效处理一维信号,但由于该方法自身的结构限制,在识别晶圆图像等二维数据时效果不佳.卷积神经网络(CNN)通过引入卷积核结构,有效解决了不同领域的图像分类和目标识别问题<sup>[20-21]</sup>,例如:GoogleNet<sup>[22]</sup>、ResNet<sup>[23]</sup>、DenseNet<sup>[24]</sup>等深度CNN在ImageNet图像库上的识别准确率已经超越人类.

在晶圆图缺陷识别领域,一些学者探究了CNN的结构,并将各种类型的晶圆图像直接作为数据输入进行分类识别<sup>[25]</sup>.也有学者将CNN运用于混合缺陷分类<sup>[26]</sup>、自动特征提取<sup>[27]</sup>、晶圆图像检索<sup>[28]</sup>等研究方向.但是当采用上述基于CNN的传统深度学习方法识别新的晶圆图缺陷诊断任务时,研究人员通常需要重新设计并训练一个新的深度卷积网络.整个过程包括复杂的超参数调节和模型参数再训练,这需要耗费大量的时间资源和计算资源<sup>[29]</sup>.迁移学习(transfer learning)<sup>[29-30]</sup>可以认为是在最小人工监督代价下进行机器学习的一种崭新策略.基于模型参数共享的迁移学习算法通过迁移在大规模数据集中预训练好的网络模型,将该模型嵌入在其他任务模型中作为特征提取器.大规模图像数据集上的实验结果证明了预训练网络提取到的特征具有良好的判别性<sup>[31]</sup>.此外,由于具有类别标签的图像不易获得,且人工采集成本较高,晶圆图数据存在着类不平衡的问题.迁移学习算法通过学习在大规模数据集上表现优异的模型和权重参数,能够有效提取小样本数据的深度特征信息并识别其类型<sup>[31]</sup>.

深度CNN提取出的特征维度较高、信息较为抽象,深度森林(Deep Forest)<sup>[32-33]</sup>这一基于随机森林算法的决策树集成深度模型采用多维度扫描和级联处理的方法,可以有效地识别并处理高维度特征信息.深度森林在随机森林的基础上,引入了多粒度扫描和级联森林2个阶段,使模型具有上下文感知能力和深度表征能力.相比于深度神经网络算法,深度森林算法具有更少的超参数,且能够自适应地决定模型的层数大小,减少模型的复杂度.实验结果表明,即使在面对不同任务时,深度森林的默认超参数设置也能取得有效的识别效果<sup>[32]</sup>.综上所述,现有晶圆图缺陷模式识别模型主要有以下3个问题.

1)传统特征提取方法配合分类器的算法作为

目前晶圆图缺陷识别的重要方法之一,其人工提取特征的方法具有一定的主观性和局限性,此外,该方法人工选取特征和分类器调参的成本较高。

2) 现有的基于深度学习的晶圆图缺陷模式识别算法虽然可以直接提取晶圆图图像特征,但整个过程往往需要重新设计整个深度神经网络架构,训练过程不易拟合,需要大量的时间资源和计算资源。深度神经网络所提取的特征往往维度较高,特征信息抽象,这给分类器的识别性能带来了挑战。

3) 从晶圆图的样本数据来看,绝大多数工业生产中的晶圆图数据集具有严重的类不平衡问题,如何有效识别数据特征不明显的小样本数据是目前的一个难点。

为此,本文提出基于迁移学习和深度森林集成的晶圆图缺陷模式识别模型 DenseNet-GCForest。DenseNet-GCForest 算法将迁移学习的思路引入深度卷积神经网络,在对数据进行标准化处理后,通过 DenseNet 网络直接从原始图像中提取深度特征,并引入 GCForest 特征分类模型,对深度特征信息进行分类识别。最后在工业数据集 WM-811K 上验证 DenseNet-GCForest 的有效性。

## 1 基于迁移学习的 DenseNet 晶圆图特征提取模型

在晶圆图缺陷模式识别过程中,采用深度卷积神经网络直接从晶圆图原始信号中提取特征。由于直接、完整地训练一个深度卷积神经网络模型会消耗大量的时间资源和计算资源,本文在 DenseNet 的基础上引入迁移学习的方法,通过迁移在大型数据集 ImageNet 上预训练好的模型权重参数,将性能优秀的模型参数初始化为本文模型的起始权重参数,并重新设计最后的全连接分类层,从而大幅度减少模型的收敛拟合时间,生成性能优异的特征提取模型。

### 1.1 迁移学习

CNN 在计算机视觉领域展现出了强大的识别性能,但这往往依赖更深的网络结构,同时意味着有更多的参数需要训练,迁移学习能够很好地解决这一问题。迁移学习将性能优秀的分类网络的前  $n$  层迁移至新的神经网络,作为新任务的特征提取器,前  $n$  层的权重参数是预训练好的,假设新的神经网络共有  $m$  层,整个网络后  $m-n$  层

需要针对新的数据集进行结构设计和权重训练。迁移学习所涉及的数据集如下式所示:

$$D_S = \{X_S, L_S\}, D_T = \{X_T, L_T\}, P_S \neq P_T. \quad (1)$$

式中:  $D_S$ 、 $D_T$  分别为旧模型和新模型的数据集,  $X$ 、 $L$  和  $P$  分别为数据信息、类标签和类分布情况,下标  $S$  和  $T$  分别为迁移学习中的源域和目标域。

源域和目标域的 CNN 分类识别过程如下式所示:

$$\hat{L}_S = M_S(X_S, w_S), \quad \hat{L}_T = M_T(X_T, w_T). \quad (2)$$

式中:  $M$  为 CNN 模型,  $\hat{L}$  为模型输出,  $w$  为模型权重参数。

预训练模型的参数训练过程如下式所示:

$$w_S = \arg \min_{w_S} [E(L_S - M_S(X_S, w_S))]. \quad (3)$$

式中:  $E$  为交叉熵损失函数。

在迁移学习过程中,预训练模型的前  $n$  层参数被迁移至新的任务中,如下式所示:

$$w_T(1:n) = w_S(1:n). \quad (4)$$

式中:  $w_T(1:n)$  和  $w_S(1:n)$  分别为源域模型和目标域模型前  $n$  层的权重参数。

目标域模型后  $m-n$  层参数将会使用新任务的数据进行训练,训练过程如下式所示:

$$w_T(1:m) = [w_T(1:n), w_T(n:m)] = \arg \min_{w_T(n:m)} \left\{ E[L_T - M_T(X_T, (w_T(1:n), w_T(n:m))) \right\}. \quad (5)$$

式中:  $w_T(1:m)$  为目标域模型的全部权重参数,在训练过程中,目标域模型的前  $n$  层的权重参数  $w_T(1:n)$  保持不变,目标域模型的后  $m-n$  层的权重参数  $w_T(n:m)$  在新数据的基础上进行训练优化。

### 1.2 DenseNet 深度卷积神经网络

DenseNet<sup>[22]</sup> 密集连接深度卷积神经网络具有强大的特征学习和特征提取能力。深度学习网络性能的提升往往伴随着网络深度的提升,但是随着网络深度的逐渐加深和模型的复杂化,深度神经网络在训练过程中的梯度消失问题会愈加严重<sup>[22]</sup>。DenseNet 全新设计了深度神经网络模型的特征传导机制,很好地解决了这个问题。DenseNet 是在 GoogLeNet<sup>[20]</sup>、ResidualNet<sup>[21]</sup> 以及 HighwayNet<sup>[34]</sup> 的基础上被提出来的。DenseNet 的核心思想是跨层连接,网络深浅层的特征信息被神经网络直接连接了起来,从而有效减轻了梯度消失的问题<sup>[22]</sup>。不同于传统神经网络过度依赖网络最后一层特征类判别性, DenseNet 可以综合利用



网络中的深浅层特征,加强了特征的传导和利用,减轻了梯度消失的问题,从而得到了具有更好泛化性能的决策函数<sup>[22]</sup>.

图 1 展示了 DenseNet 的主要结构——密集连接模块(Dense block)的示意图. Dense block 中每一层的变换如下式所示:

$$x_l = H_l\{[x_0, x_1, \cdots, x_{l-1}]\}.$$
 (6)

式中:  $x_l$  为第  $l$  层网络的输出,  $H_l$  为第  $l$  层的输出函数表示, 每一层网络包含了批标准化(batch normalization, BN)、Relu 函数和卷积操作<sup>[22]</sup>.

在 DenseNet 网络中,  $H_4$  层的输入不仅包含了上一层  $H_3$  的直接输出, 同时包括了原始信息  $x_0$  以及  $H_1$ 、 $H_2$  层中的全部特征信息. DenseNet 在反向

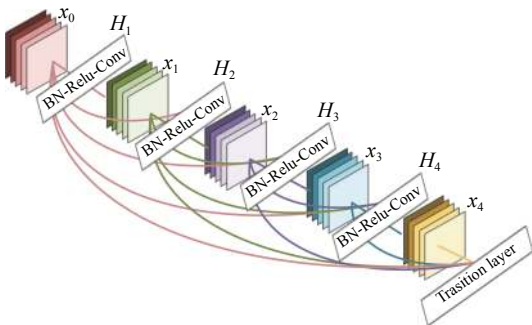


图 1 Dense Block 密集连接模块框架

Fig.1 Dense block connection module framework

传播的训练过程中,  $x_l$  的梯度信息包含了损失函数直接对 Dense block 中每一层的输出信息计算得到的导数.

1.3 基于迁移学习的 DenseNet169

DenseNet 一般由多个密集连接模块和过渡层组成, 本文通过迁移学习的方法, 以网络结构深度为 169 的密集连接神经网络 (DenseNet169) 为基础, 迁移 DenseNet169 在 ImageNet 数据集上预训练好的前 4 个密集连接模块, 并重新设计分类神经网络结构. 由于本实验共有 9 个类别的晶圆图片, 在晶圆图缺陷模式识别任务中使用 9 维全连接层代替原有分类层中的 1000 维全连接层. 图 2 展示了晶圆图识别任务中基于迁移学习的 DenseNet169 网络结构和模型参数, 新网络由迁移模型和新任务模型两部分组成. 结构表中  $k$  代表每个卷积层滤波器的数量, 每个密集连接模块中交替使用  $1\times 1$  和  $3\times 3$  的卷积操作. 分类层中采用 Softmax 函数将神经元输出转化为概率:

$$\text{Softmax} \{[x_1, x_2, \cdots, x_n]^T\} = \left[ \frac{\exp(x_1)}{\sum_{i=1}^n \exp(x_i)}, \frac{\exp(x_2)}{\sum_{i=1}^n \exp(x_i)}, \cdots, \frac{\exp(x_n)}{\sum_{i=1}^n \exp(x_i)} \right]^T.$$
 (7)

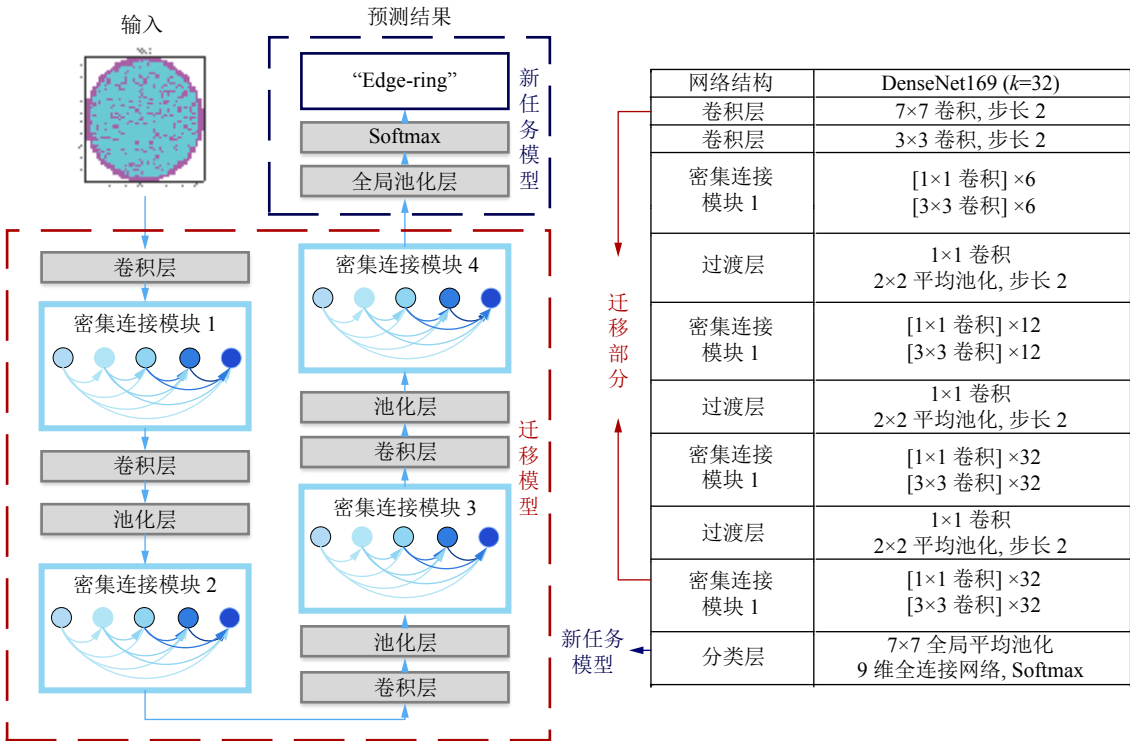


图 2 基于迁移学习的 DenseNet 网络结构

Fig.2 DenseNet network structure based on transfer learning

式中:  $x_1, x_2, \dots, x_n$  为神经网络全连接层的输出值, 通过 Softmax 函数将对应的值映射到  $[0, 1.0]$ , 将预测值转化为可能事件的概率分布.

DenseNet169 采用交叉熵损失函数作为目标函数来衡量识别结果和真实类别的误差, 其损失为

$$H(p, q) = - \sum_x p(x) \log q(x). \quad (8)$$

式中:  $p$  与  $q$  分别为真实分布和预测分布, 对应本模型中的晶圆图像实际类别和模型预测分布.

DenseNet169 采用随机梯度下降算法对网络进行训练:

$$w_i = w_i - \alpha [h_\theta(x_0^j, x_1^j, \dots, x_n^j) - y_j] x_i. \quad (9)$$

式中:  $j$  为网络中第  $j$  层,  $i$  为每层中的神经元,  $w_i$  为需要学习的权重参数,  $\alpha$  为当前学习率,  $h_\theta$  为神经网络的预测运算.

整个 DenseNet 网络在训练过程中冻结前 3 个 DenseBlock, 并对后面的网络进行局部微调, 当模型收敛后, 将最后分类层中全局池化操作所得到的数据特征抽取出来, 作为晶圆图的深度特征信息.

## 2 多粒度扫描深度森林特征分类模型

通过 DenseNet169 提取出的特征维度较高, 特征信息抽象, 直接使用全连接神经网络并不能有效识别各个晶圆图的类别. 深度森林 (Deep forest) 这一基于随机森林算法的决策树深度模型采用多维度扫描和级联处理的方法, 可以有效地识别和处理高维度特征信息, 并在模式识别问题上表现出了良好的性能. 多维度扫描集成森林 GCForest 是一种新型决策树集成方法<sup>[32,35-36]</sup>, 该方法在随机森林的基础上引入了级联结构和集成学习, 从而增强了决策树模型的表征能力. 当输入是高维数据时, GCForest 通过多维度扫描生成多尺度特征向量, 从而进一步增强算法的表征学习能力. GCForest 可以自适应地确定级联层数, 自行决定模型复杂程度, 用户可以根据可用的计算资源来控制训练成本. GCForest 的另一优点在于该模型具有较少的超参数, 且超参数具有较强的鲁棒性.

GCForest 由 2 个部分所组成. 第一部分是多粒度扫描结构, GCForest 通过滑动窗口扫描原始数据, 生成一组多尺寸特征向量. 第二部分是级联森林结构, 其重要思想之一是根据每个输入实例生成对应类分布. 如图 3 所示, GCForest 通过计算决策树中各叶节点处不同类实例数量占比的平

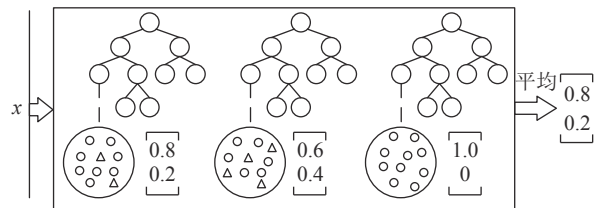


图 3 GCForest 的类向量生成

Fig.3 Class vector generation in GCForest

均值, 从而预测各实例的类别分布. 图 4 展示了深度森林的网络结构, 多粒度扫描阶段首先根据输入数据生成多尺度特征向量. 其次, 级联森林<sup>[32]</sup>阶段对生成的多尺度特征向量进行逐层训练, 预测其类别分布, 并将生成的类别分布向量与原始向量连接, 通过级联通道以输入到下级级联层. 在级联森林的训练过程中, 当验证集的识别性能收敛时, 模型自适应终止训练, 因此, 级联森林能够自适应地决定模型复杂度. 在本研究的级联森林中, 每一层由 2 个随机森林和 2 个完全随机森林组成. 其中, 每个随机森林包含 500 棵决策树, 每个完全随机森林则包含 1 000 棵决策树, 两者的区别主要在于候选特征空间的选择. 完全随机森林在完整的特征空间中随机选择特征作为决策树的分裂节点, 然后不断生长, 当决策树中的所有叶节点都无法分裂时, 则终止训练; 普通随机森林随机选取  $\sqrt{d}$  ( $d$  为输入特征维度) 个候选特征, 然后通过 Gini 指数筛选分裂节点进行生长分裂<sup>[35]</sup>. 基尼指数公式如下:

$$\text{Gini}(p) = \sum_{k=1}^K p_k(1-p_k) = 1 - \sum_{k=1}^K p_k^2. \quad (10)$$

式中: 假设样本的可能取值为  $\{1, 2, \dots, K\}$ ,  $p_k$  为样本属于  $k$  类别的概率.

## 3 DenseNet-GCForest 晶圆图缺陷模式识别模型

### 3.1 DenseNet-GCForest

DenseNet-GCForest 首先将 DenseNet169 在 ImageNet 上训练的网络参数作为本模型的初始参数, 并舍弃模型的全局池化层和分类层, 重新设计全连接层和输出层. 在冻结网络前 3 个 DenseBlock 参数的情况下, 采用反向传播和随机梯度下降算法微调剩余局部网络权重参数, 训练得到一个晶圆图像特征提取模型, 抽取最后全局池化层的数据作为晶圆图的高维抽象特征. 采用深度森林算法, 通过滑动窗口和级联森林进一步对所提

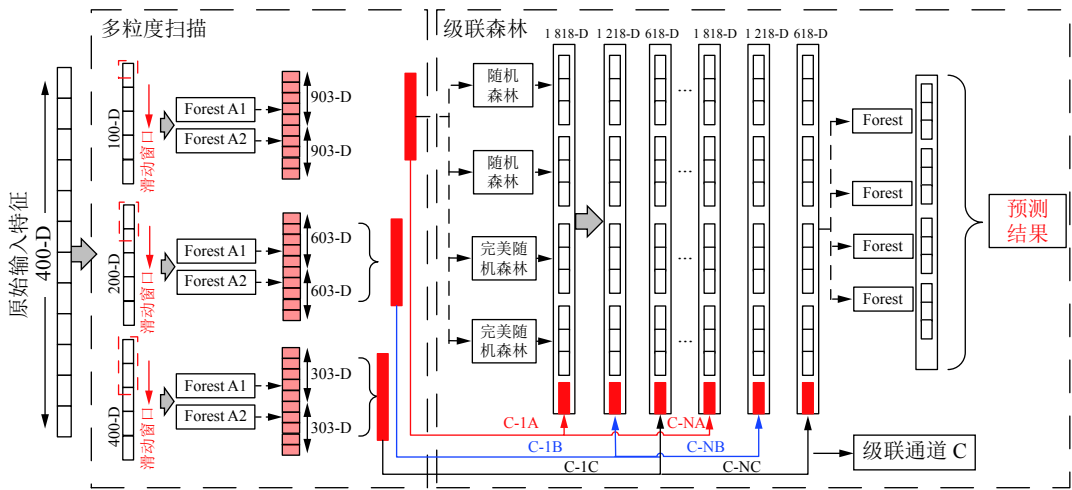


图 4 深度森林网络结构 (以 400 维输入为例)

Fig.4 Deep forest network structure ( taking 400-dimensional input as example )

取的高维图像特征进行感知分类,从而得到最终的预测结果.

3.2 DenseNet-GCForest 应用流程

如图 5 所示为 DenseNet-GCForest 晶圆图缺陷识别算法的应用流程, DenseNet-GCForest 算法分为离线建模与在线识别两部分,具体流程如下.

- 1) 离线建模:
  - a. 读取晶圆图像并按照一定比例将其随机分为训练集和测试集;
  - b. 迁移 ImageNet 数据集上预训练好的 DenseNet169 模型及权重参数;
  - c. 重新设计全连接层和池化层并通过训练集进行微调训练,抽取全局平均池化层作为晶圆图像特征,形成特征提取模型;
  - d. 将训练集图片输入特征提取模型,得到晶

- 圆图像高维特征;
- e. 在训练集特征上训练深度森林模型.
- 2) 在线检测:
  - a. 在线采集晶圆图像,将图像输入至 DenseNet169 模型进行特征提取;
  - b. 通过 DenseNet169 提取晶圆图缺陷的高维抽象特征;
  - c. 将提取的高维抽象特征模型输入至 GCForest 模型进行预测;
  - d. 得到最终的识别分类结果并进行验证.

4 实验与结果分析

利用现场采集的工业数据集 WM-811K<sup>[37]</sup> 晶圆图缺陷样本库来检验算法对缺陷的识别能力,该样本库中的图像均来自于实际半导体生产线.该样本库中浅灰色、深灰色和白色像素分别表示每个晶片图的正常、有缺陷和空元素部分. WM-811K 数据集由正常模式和 8 种缺陷模式组成,图 6 展现了中心 (Center)、圆环 (Donut)、边缘局部 (Edge-local)、边缘环 (Edge-ring)、局部 (Local)、近满 (Near-full)、随机 (Random),划痕缺陷 (Scratch) 和正常无缺陷 (None-pattern) 共 9 种类型的图案.图 7 列出了 WM-811K 数据集的各类缺陷图像的数量  $N$ ,由此可见,该样本库存在明显的类不平衡,这对正确识别缺陷模式带来了很大挑战.每个晶圆图的大小被预处理为  $128 \times 128$  像素.本实验将这些图像随机分成 5 份,进行五折交叉验证.每次验证随机选取 7 112 张晶圆图作为训练数据集,剩余 2 000 张作为验证数据集.

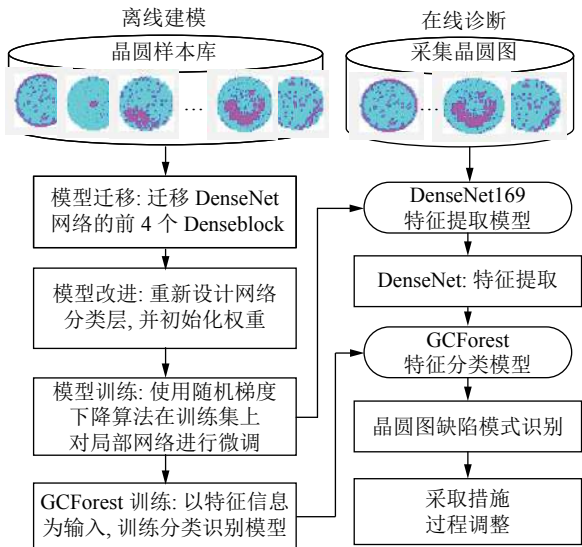


图 5 DenseNet-GCForest 晶圆图缺陷识别流程

Fig.5 Defect identification process of DenseNet-GCForest wafer map



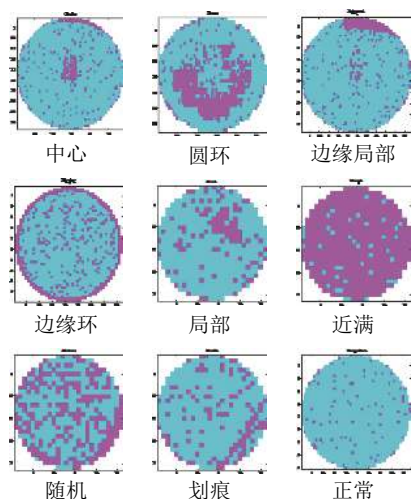


图 6 正常晶圆图模式与 8 种缺陷模式

Fig.6 Normal wafer maps and 8 defect modes

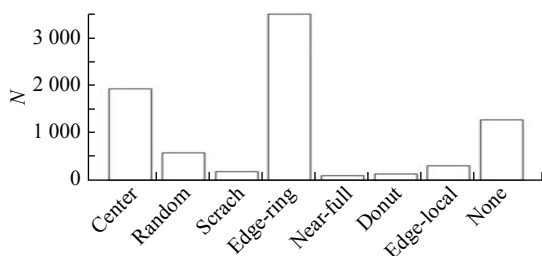


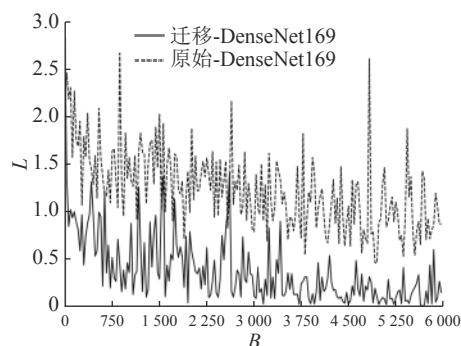
图 7 晶圆缺陷类型数量统计

Fig.7 Number of different types of wafer maps

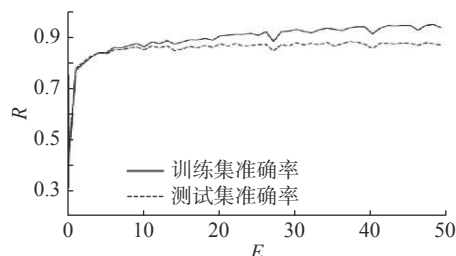
#### 4.1 基于迁移学习的 DenseNet 特征学习

本文采用基于迁移学习的 DenseNet169 提取晶圆图像特征,为了验证迁移学习的有效性,测试原始 DenseNet169 和基于迁移学习的 DenseNet169 神经网络的训练过程.图 8(a)记录了原始 DenseNet169 和基于迁移学习的 DenseNet169 在训练过程中损失函数的变化曲线,其中,  $B$  为训练过程中的批次数,  $L$  为模型的交叉熵损失函数值.从图中可知,基于迁移学习的 DenseNet169 通过不到 3 000 批次迭代训练,模型基本达到收敛,相比于直接对整个网路进行训练,基于迁移学习的网络损失函数值更小、收敛速度更快.图 8(b)记录了基于迁移学习的 DenseNet169 在训练数据集和测试数据集上的识别率变化曲线.图中,  $E$  为训练轮数,  $R$  为识别准确率.从图中可知,直接使用反向传播神经网络进行识别分类,最终识别准确率不到 80%,因此本文进一步引入深度森林模型对高维抽象数据进行分类,进而提高模型的识别性能.

图 9 为该晶圆图在 DenseNet 网络中第二、三、四卷积模块的特征输出图,随着层数的递增,特征逐渐细化、抽象化,整个神经网络对特征的



(a) 损失函数曲线



(b) 识别率曲线

图 8 基于迁移学习的 DenseNet169 的训练过程

Fig.8 Training process of DenseNet169 based on transfer learning

关注重点由整体转入局部细节.从第四卷积层的输出中,已经无法分辨出图像的特点,这充分表明了 DenseNet 后面卷积层的输出特征与前面相比抽象程度更高.

为了探究 DenseNet 网络提取特征的有效性,运用 T-SNE<sup>[38]</sup> 数据可视化算法,对原始图像和 DenseNet 网络的输出特征进行数据降维,将不同类别的高维原始图像和特征信息降维至二维平面,并运用散点图描绘相关数据信息.图 10 展示了 T-SNE 数据可视化的结果,在二维空间 (D1D2) 中,原始图像的多个类别混合在一起,极难判别,而 DenseNet 网络提取的特征有效分离了 9 个晶圆图缺陷类型,该特征具有良好的类判别性,说明 DenseNet169 具有优秀的特征提取能力.

#### 4.2 识别性能分析

如表 1 所示为 DenseNet-GCForest 晶圆缺陷识别率混淆矩阵,由表 1 可知, DenseNet-GCForest 在测试数据集上的综合识别准确率达到 96.2%.由各个类别的识别率可知,除了随机型缺陷外,本文模型能够有效识别其他类型的晶圆图类别,识别准确率都在 93.75% 以上.除此之外,如图 7 所示的晶圆图数据集有着明显的类不平衡问题,近满、局部和划痕型晶圆图的数据量较少,而 DenseNet-GCForest 通过采用迁移学习和深度森林的方法,能够有效学习这类晶圆图的数据特征,获得了非常好的识别效果.

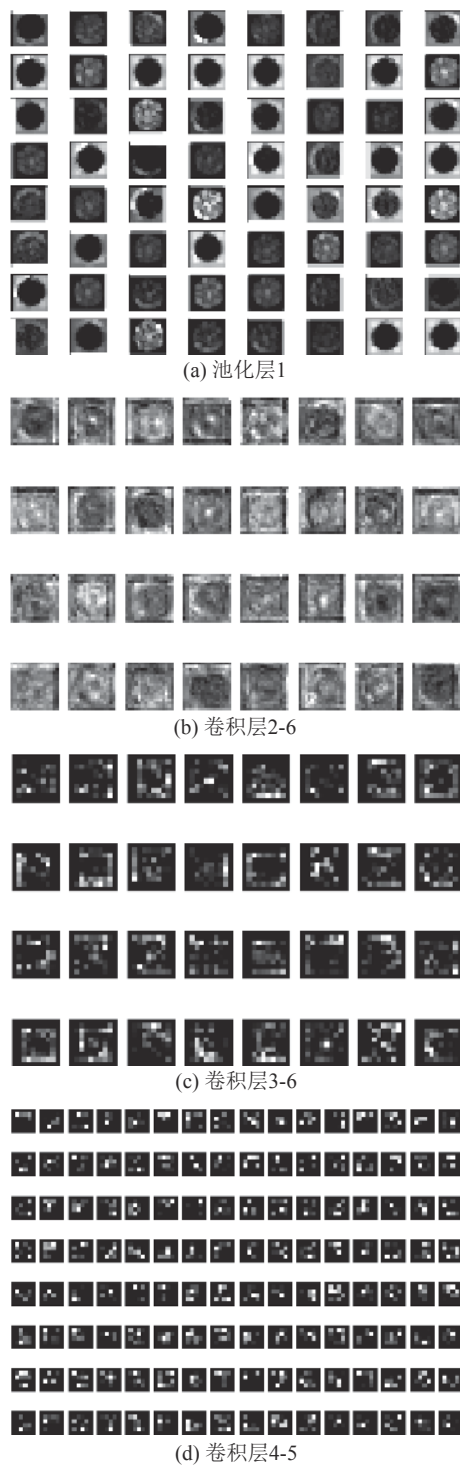


图 9 晶圆图多级网络输出特征

Fig.9 Output features of wafer maps in multilevel networks

对于 Random 类晶圆图缺陷, 进一步分析其误识别原因. 由表 1 可知, 有 5.13%、2.56%、2.56% 和 2.56% 的 Random 晶圆图被误识别成了 Local、Near\_full、Edge\_local、Center 和 Edge\_ring 类. 如图 11 所示为 4 张被误识别为其他类别的 Random 晶圆图, 可知, 其在具有标准随机缺陷的同时, 依次还具有 Local、Near\_full、Center 和 Edge\_ring 的缺陷特征(见图 6), 从而被模型误判.

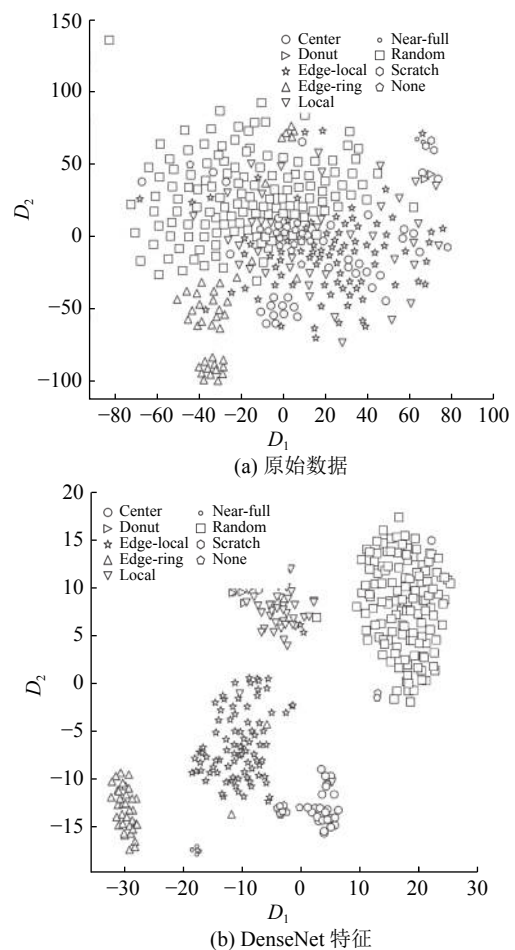


图 10 原始数据和 DenseNet 特征的可视化分析

Fig.10 Visualization analysis of raw data and DenseNet features

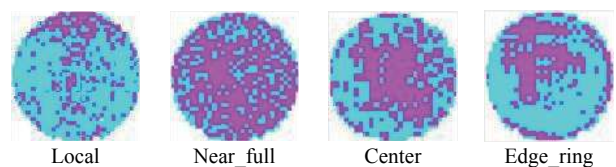


图 11 Random 类的误识别情况

Fig.11 Error identification of Random class

### 4.3 参数灵敏度分析

多粒度扫描森林的主要超参数分为 2 类: 一是结构参数; 二是随机森林生长规则参数, 其主要超参数有森林数数量、每个森林的决策树数量、滑动窗口大小、级联的每层森林数量、决策树停止生长规则. 本文主要针对随机森林生成规则参数进行灵敏度分析, 如表 2 所示, 其中,  $R_{acc}$  为预测准确率, 表示预测为正的样本中真正的正样本所占的比例. 从表 2 可以看出, 在其他参数一样的情况下, 单独调整某一参数对实验结果的影响不大. 从实验结果可知, GCForest 的性能对于超参数的设置具有很强的鲁棒性, DenseNet-GC-



表 1 Densenet-GCForest 晶圆缺陷识别率混淆矩阵

Tab.1 Confusion matrix of Densenet-GCForest wafer map defect recognition rate

%

预测真实	None	Center	Donut	Edge-local	Edge-ring	Local	Near-full	Random	Scratch
None	99.59	0.00	0	0.28	0	0	0	0	0.14
Center	0.83	97.52	0	0.83	0	0	0	0	0.83
Donut	0	0	94.74	0	0	5.26	0	0	0
Edge-local	0.86	0.29	0	96.83	0.29	1.44	0	0.29	0
Edge-ring	0	0	0	4.91	95.09	0	0	0	0
Local	0.75	0.38	0	1.51	0	96.23	0	0.75	0.38
Near-full	0	0	0	0	0	0	94.12	5.88	0
Random	0	2.56	0	0	2.56	5.13	2.56	87.18	0
Scratch	1.25	0	0	1.25	0	3.75	0	0	93.75

表 2 GCForest 识别器参数灵敏度分析

Tab.2 Parameter sensitivity analysis of GCForest

参数名称	参数大小	$R_{acc}/\%$
滑动窗口大小	400	97.31
	700	96.7
	1 000	96.2
	1 300	96.4
	1 664	96.2
决策树生成的 最小样本数	0	96.2
	0.1	96.7
	0.2	96.2
	0.3	95.8
	0.4	95.4
决策树生成的 允许误差	0	96.2
	0.1	96.7
	0.2	96.2
	0.3	95.8
	0.4	95.4
扫描层随机森林的 决策树数量	200	96.9
	400	95.8
	600	97.0
	800	96.7
	1 000	96.0

Forest 模型训练便捷,且训练的时间成本少于一般深度学习方法.

4.4 性能比较

为了验证所提方法的有效性和先进性,实验

中将几种经典和最新的晶圆图缺陷模式识别方法作为 DenseNet-GCForest 的比较对象. 其中, CNN+Softmax 分类器 (CNN-Softmax) 方法的 CNN 模型选择性能优秀的 GoogleNet<sup>[15]</sup>、ResNet<sup>[14]</sup> 和 DenseNet<sup>[16]</sup>. 将决策树 C4.5 的最大深度设置为 25, 节点数目为 100. 支持向量机分别采用线性核函数和高斯核函数, 将惩罚因子设置为  $C=1.0$ . 将随机森林最大深度设置为 50, 包含 800 棵树. 将 KNN 分类器中的  $K$  值设置为 5. 决策树集成算法<sup>[14]</sup> 将 4 种根据不同特征 (最大值、最小值、均值和方差) 形成的分类器进行集成投票. 堆叠降噪自编码器 (stacked denoising auto encoder, SDAE)<sup>[28]</sup> 深度学习算法共包含 2 个隐层: 第一隐层由 100 个神经元组成, 第二隐层包含 50 个神经元.

如表 3 所示为 DenseNet-GCForest 与经典 CNNs、决策树集成和 SDAE 在各类晶圆图缺陷上的识别性能对比. 其中,  $R_{rec}$  为召回率, 表示在所有正样本中被预测正确的正样本所占比例;  $F$  为两者的调和平均数, 用于表示模型识别性能的稳定性,  $F=2R_{acc}R_{rec}/(R_{acc}+R_{rec})$ . 从表中可知, DenseNet-GCForest 算法在 Center、Donut、Edge\_ring、Edge\_local、Local、Scratch 这 6 种晶圆图缺陷模式识别任务中, 展现出了最佳的性能. 即使对于另外 Near\_full、None、Random 3 种缺陷模式, 其  $F$  值也达到了 94.1、99.3 和 88.3. 对于这 3 种缺陷类型, 所有方法中的最高  $F$  值分别为 95.7、99.7、91.2, 本文算法的识别性能接近这些最优  $F$  值. 其中, Near\_full 和 None 这 2 类缺陷的特征非常明显, 采用各个识别算法都能对其进行有效识别. 此外, 所提出算法对于缺陷特征不够明显且样本数量较

表 3 DenseNet-GCForest 算法的对比实验结果

Tab.3 Performance comparison of DenseNet-GCForest and other algorithms

%				
实验方法	缺陷类别	$R_{acc}$	$R_{rec}$	$F$
DenseNet-GCForest	Center	97.5	97.5	<b>97.5</b>
	Donut	100	94.7	<b>97.3</b>
	Edge_local	95.5	96.8	<b>96.1</b>
	Edge_ring	98.7	95.1	<b>96.9</b>
	Local	95.9	96.2	<b>96.0</b>
	Near_full	94.1	94.1	94.1
	None	99.0	99.6	99.3
	Random	89.5	87.2	88.3
	Scratch	96.2	93.8	<b>94.9</b>
GoogleNet	Center	72.2	34.5	46.7
	Donut	59.1	43.3	50.0
	Edge_local	63.9	79.8	71.0
	Edge_ring	93.4	87.6	90.4
	Local	57.3	48.8	52.7
	Near_full	93.3	93.3	93.3
	None	90.6	92.9	91.7
	Random	85.4	92.1	88.6
	Scratch	81.7	81.7	81.7
ResNet	Center	85.0	77.1	80.9
	Donut	47.2	81.0	59.6
	Edge_local	87.6	81.4	84.4
	Edge_ring	94.3	86.6	90.3
	Local	81.7	68.6	74.6
	Near_full	75.0	75.0	75.0
	None	89.9	99.3	94.4
	Random	88.2	73.2	80.0
	Scratch	59.5	71.0	64.4
DenseNet	Center	76.4	85.5	80.7
	Donut	95.2	64.5	76.9
	Edge_local	79.7	91.2	85.0
	Edge_ring	96.1	81.5	88.2
	Local	82.1	66.8	73.6
	Near_full	91.7	100	<b>95.7</b>
	None	95.2	99.3	97.2
	Random	86.4	65.5	74.5
	Scratch	81.5	72.6	76.8

续表 3

实验方法	缺陷类别	$R_{acc}$	$R_{rec}$	$F$
Decision TreeEnsemble	Center	95.6	93.75	94.7
	Donut	92.6	92.3	92.4
	Edge_local	83.5	87.3	85.4
	Edge_ring	86.8	91.1	88.9
	Local	83.5	82.3	82.9
	Near_full	89.4	91.7	90.5
	None	100	99.5	<b>99.7</b>
	Random	91.7	87.3	88.4
	Scratch	86.0	88.5	87.2
SDAE	Center	98.5	87.3	92.6
	Donut	87.6	89.1	88.4
	Edge_local	87.5	85.4	86.4
	Edge_ring	98.8	94.7	96.7
	Local	83.5	55.8	66.9
	Near_full	85.3	96.2	90.4
	None	99.0	98.7	98.9
	Random	93.3	89.3	<b>91.2</b>
	Scratch	81.7	39.6	53.3

少的 Donut 和 Scratch 缺陷模式的识别准确率为 97.3% 和 94.9%, 明显优于对比算法 (5% ~ 40%), 因此认为基于迁移学习的 DenseNet-GCForest 晶圆图缺陷识别对小样本数据有较好的识别能力.

为了避免随机因素对实验结果的影响, 验证所提出的模型性能的稳定性, 进行五折交叉验证. 基于五折交叉验证的不同识别器的平均识别率如表 4 所示. 可知, DenseNet-GCForest 的识别效果非常突出, 显著优于所有其他识别器. 此外, 当将 DenseNet 所提取的特征应用到所有的分类算法时,

表 4 五折交叉识别率对比

Tab.4 Comparison of five-fold cross validation of various algorithms

分类器	$R_{acc}/\%$	分类器	$R_{acc}/\%$
DenseNet-GCForest	<b>96.8</b>	GCForest	<b>73.7</b>
C4.5 Ensemble	90.8	RF	68.9
SDAE	89.4	SVML	72.5
DenseNet	88.6	SVMG	40.2
GoogleNet	74.3	KNN	30.1
ResNet	86.5	C4.5	62.4

几乎所有算法的识别准确率都得到了提高,如表5所示,DenseNet提取的特征提高了所有典型识别器的性能,GCForest仍然具有更好的识别性.

表 5 基于 DenseNet 特征的多种识别器五折交叉识别率对比  
Tab.5 Comparison of five-fold cross recognition rate of various recognizers based on DenseNet features

分类器	$R_{acc}/\%$	分类器	$R_{acc}/\%$
GCForest	96.8	SVMG	95.5
BPN	85.1	KNN	92.6
RF	95.1	C4.5	87.2
SVML	95.6	—	—

5 结 语

本文提出了基于迁移学习与深度森林的晶圆图缺陷模式识别模型 DenseNet-GCForest,在实际工业数据集上的实验结果表明,该方法的识别准确率达到 了 96.8%,其性能优于典型的卷积神经网络以及其他常用识别方法.该方法主要具有以下优点:

- 1)通过深度卷积神经网络直接从晶圆缺陷原始图像中提取特征,能够有效避免人工选取特征的主观性和局限性,同时提取的特征具有良好的分类特性;
- 2)通过迁移预训练好的特征提取模型,在保证特征提取的有效性上极大减少了模型训练时间并提升了模型的图像特征提取能力;
- 3)引入了深度森林算法对高维抽象特征进行分类,该方法作为基于随机森林的集成方法,吸收了深度神经网络的思想,具有良好的处理表征关系的能力和逐层加强表征学习的能力,相比于传统算法具有更好的分类性能.

参考文献 (References):

[1] TAO Y, WAY K, SUK J B. Detection of spatial defect patterns generated in semiconductor fabrication processes [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2011, 24(3): 392–403.

[2] KIM B, JEONG Y S, TONG S H, et al. Step-down spatial randomness test for detecting abnormalities in DRAM wafers with multiple spatial maps [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2016, 29(1): 57–65.

[3] YU J, LU X. Wafer map defect detection and recognition using joint local and nonlocal linear discriminant analysis [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2016, 29(1): 33–42.

[4] SHANKAR N G, ZHONG Z W. Defect detection on semiconductor wafer surfaces [J]. **Microelectronic Engineering**, 2005, 77(3): 337–346.

[5] LEE S H, KOO H I, CHO N I. New automatic defect classification algorithm based on a classification-after-segmentation framework [J]. **Journal of Electronic Imaging**, 2010, 19(2): 334–343.

[6] HWANG J Y, KUO W. Model-based clustering for integrated circuit yield enhancement [J]. **European Journal of Operational Research**, 2007, 178(1): 143–153.

[7] BALY R, HAJJ H. Wafer classification using support vector machines [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2012, 25(3): 373–383.

[8] XIE L, HUANG R, GU N, et al. A novel defect detection and identification method in optical inspection [J]. **Neural Computing and Applications**, 2014, 24(7/8): 1953–1962.

[9] CHAO L C, TONG L I. Wafer defect pattern recognition by multi-class support vector machines by using a novel defect cluster index [J]. **Expert Systems with Applications**, 2009, 36(6): 10158–10167.

[10] ADLY F, ALUSSEIN O, YOO P, et al. Simplified subspace regression network for identification of defect patterns in semiconductor wafer maps [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2015, 11(6): 1267–1276.

[11] KIM B, JEONG Y S, TONG S H, et al. A regularized singular value decomposition-based approach for failure pattern classification on fail bit map in a DRAM wafer [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2015, 28(1): 41–49.

[12] OOI M P L, SOK H K, KUANG Y C, et al. Defect cluster recognition system for fabricated semiconductor wafers [J]. **Engineering Applications of Artificial Intelligence**, 2013, 26(3): 1029–1043.

[13] 余建波, 卢笑蕾, 宗卫周. 基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别 [J]. **自动化学报**, 2016, 42(1): 47–59.

YU Jian-bo, LU Xiao-lei, ZONG Wei-zhou. Wafer defect detection and recognition based on local and nonlocal linear discriminant analysis and dynamic ensemble of gaussian mixture models [J]. **ACTA Automatica Science**, 2016, 42(1): 47–59.

[14] PIAO M, JIN C H, LEE J Y, et al. Decision tree ensemble-based wafer map failure pattern recognition based on radon transform-based features [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2018, 31(2): 250–257.

[15] SAQLAIN M, JARGALSAIKHAN B, LEE J Y. A voting



- ensemble classifier for wafer map defect patterns identification in semiconductor Manufacturing [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2019, 32(2): 171–182.
- [16] LECUN Y, BENGIO Y, HINTON G. Deep learning [J]. **Nature**, 2015, 521: 436–444.
- [17] 杨婧, 耿辰, 王海林, 纪建松, 等. 基于 DenseNet 的低分辨 CT 影像肺腺癌组织学亚型分类 [J]. 浙江大学学报: 工学版, 2019, 53(6): 1164–1170.
- YANG Jing, GENG Chen, WANG Hai-lin, et al. Classification on histological subtypes of lung adenocarcinoma from low-resolution CT images based on DenseNet [J]. **Journal of Zhejiang University: Engineering Science**, 2019, 53(6): 1164–1170.
- [18] THIRUKOVALURU R, DIXIT S, SEVAKULA RK et al. Generating feature sets for fault diagnosis using denoising stacked auto-encoder [C] // **IEEE International Conference in Prognostics and Health Management (ICPHM)**. Ottawa: IEEE, 2016: 1–7.
- [19] JIANG G, HE H, XIE P, et al. Stacked Multilevel-Denoising Autoencoders: a new representation learning approach for wind turbine gearbox fault diagnosis [J]. **IEEE Transactions on Instrumentation and Measurement**, 2017, 66: 2391–2402.
- [20] KRIZHEVSKY A, SUTSKEVER I, HINTON GE. Imagenet classification with deep convolutional neural networks [C] // **Advances in Neural Information Processing Systems (NIPS)**. Lake Tahoe: [s. n.], 2012: 1097–1105.
- [21] 袁公萍, 汤一平, 韩旺明, 等. 基于深度卷积神经网络的车型识别方法 [J]. 浙江大学学报: 工学版, 2018, 52(4): 694–702.
- YUAN Gong-ping, TANG Yi-ping, et al. Vehicle category recognition based on deep convolutional neural network [J]. **Journal of Zhejiang University: Engineering Science**, 2018, 52(4): 694–702.
- [22] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // **2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Boston: IEEE, 2015: 1–9.
- [23] 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: IEEE, 2016: 770–778..
- [24] HUANG G, LIU Z, MAATEN L V D et al. Densely connected convolutional networks [C] // **2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**. Honolulu IEEE, 2017: 2261–2260.
- [25] DANIEL W, BERND S R, MOSHE S. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection [J]. **CIRP Annual Manufacturing Technology**, 2016, 65(1): 417–420.
- [26] NAKAZAWA T, KULKARNI D V. Wafer map defect pattern classification and image retrieval using convolutional neural network [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2018, 31(2): 309–314.
- [27] KYEONG K, KIM H. Classification of mixed-type defect patterns in wafer bin maps using convolutional neural networks [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2018, 31(3): 395–402.
- [28] LEE H, KIM Y, KIM C O. A deep learning model for robust wafer fault monitoring with sensor measurement noise [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2017, 30(1): 23–31.
- [29] PAN S J, YANG Q. A survey on transfer learning [J]. **IEEE Transactions on Knowledge and Data Engineering**, 2010, 22(10): 1345–1359.
- [30] 张雪松, 庄严, 闫飞, 等. 基于迁移学习的类别级物体识别与检测研究与进展 [J]. 自动化学报, 2019, 45(7): 1224–1243.
- ZHANG Xue-song, ZHUANG Yan, YAN Fei, et al. Status and development of transfer learning based category-Level object recognition and detection [J]. **Acta Automatica Sinica**, 2019, 45(7): 1224–1243.
- [31] DONAHUE J, JIA Y, VINYALS O et al. DeCAF: a deep convolutional activation feature for generic visual recognition [C] // **International Conference on Machine Learning (ICML)**. Bengjin: [s. n.], 2014: 647–655.
- [32] ZHOU Z H, FENG J. Deep forest: towards an alternative to deep neural networks [C] // **International Joint Conference on Artificial Intelligence (IJCAI)**. Melbourne: [s. n.], 2017: 3553–3559.
- [33] UTKIN L V, RYABININ M A. A deep Forest for transductive transfer learning by using a consensus measure [C] // **Conference on Artificial Intelligence and Natural Language (AINL)**. Petersburg: Springer, 2017: 194–208.
- [34] SRIVASTAVA R K, GREFF K, SCHMIDHUBER J, et al. Highway networks [EB/OL]. (2015-11-03) [2019-04-30], <https://arxiv.org/abs/1505.00387>
- [35] LIU F T, TING K M, YU Y, et al. Spectrum of variable-random trees [J]. **Journal of Artificial Intelligence Research**, 2008, 32(1): 355–384.
- [36] GUO Y C. Knowledge-enabled short-term load forecasting based on pattern-base using classification & regression tree and support vector regression [C] // **Fifth International Conference on Natural Computation**. Tianjin: IEEE, 2009: 425–429.
- [37] WU M J, JANG J S R, CHEN J L. Wafer map failure pattern recognition and similarity ranking for large-scale data sets [J]. **IEEE Transactions on Semiconductor Manufacturing**, 2015, 28(1): 1–12.
- [38] LAURENS V D M, GEOFFREY H. 2008. Visualizing data using T-SNE [J]. **Journal of Machine Learning Research**, 2008, 9(11): 2578–2605.