文章编号　1004-924X(2023)20-3065-12

基于知识蒸馏的低分辨率陶瓷基板图像瑕疵检测

郭　峰[1], 孙小栋[1], 朱启兵[1*], 黄　敏[1], 徐晓祥[2]
(1. 江南大学 轻工过程先进控制教育部重点实验室,江苏 无锡 214122;
2. 无锡市创凯电气控制设备有限公司,江苏 无锡 214400)

摘要:陶瓷基板是电子器件的重要基础材料,利用机器视觉技术结合深度学习策略实现陶瓷基板的瑕疵检测对保证产品质量具有重要的意义。增加成像设备的视场以实现多个陶瓷基板的同时成像,可以显著提高陶瓷基板的检测速度;但也带来了图像分辨率的降低,并最终导致瑕疵检测精度的降低。针对上述问题,本文提出了一种基于知识蒸馏的低分辨率陶瓷基板瑕疵自动检测方法。该方法利用YOLOv5框架分别构建了教师网络和学生网络,基于知识蒸馏思想将教师网络获得的高分辨率图像特征信息指导学生网络的训练,以提高学生网络对低分辨率陶瓷基板图像的瑕疵检测能力;同时,在教师网络中引入基于Coordinate Attention (CA)注意力思想的特征融合模块,使得教师网络学习到的特征同时适应高分辨率图像信息和低分辨率图像信息,从而能较好地指导学生网络的训练;最后,引入基于Gradient Harmonizing Mechanism(GHM)的置信度损失函数,以提高瑕疵的检出率。实验结果表明,本文基于知识蒸馏的陶瓷基板瑕疵检测方法对于224×224分辨率输入图像的污渍、异物、多金、缺瓷以及损伤这五类瑕疵检测的平均准确率和平均召回率分别达到了96.80%和90.01%,相比于目前主流的目标检测算法,本文算法取得了更好的检测结果。

关　键　词:陶瓷基板;瑕疵检测;YOLOv5;知识蒸馏
中图分类号:TP391.4　　**文献标识码:**A　　**doi:**10.37188/OPE.20233120.3065

Defect detection of low-resolution ceramic substrate image based on knowledge distillation

GUO Feng[1], SUN Xiaodong[1], ZHU Qibing[1*], HUANG Min[1], XU Xiaoxiang[2]

(1. *Key Laboratory of Advanced Process Control for Light Industry*, *Ministry of Education*, *Jiangnan University*, *Wuxi* 214122, *China*;
2. *Wuxi CK Electric Control Equipment Co.*, *Ltd*, *Wuxi* 214400, *China*)
* *Corresponding author*, *E-mail*: *zhuqib*@163. *com*.

Abstract: Ceramic substrate is a vital foundational material of electronic devices, and implementing defect detection for ceramic substrates using machine vision technology combined with deep learning strategies holds significant importance in ensuring product quality. Increasing the field of view of the imaging equipment to make simultaneous imaging of multiple ceramic substrates possible can significantly improve the detection speed of a ceramic substrate. However, it also results in decreased image resolution and subsequently reduces the accuracy of defect detection. To solve these problems, a low-resolution ceramic substrate defect automatic detection method based on knowledge distillation is proposed. The method utilizes

收稿日期:2023-04-10;**修订日期:**2023-07-17.
基金项目:国家自然科学基金资助项目(No. 62273166)

the YOLOv5 framework to construct a teacher network and a student network. Based on the idea of knowledge distillation, high-resolution image feature information obtained by the teacher network is used to guide the training of the student network to improve the defect detection ability of the student network for low-resolution ceramic substrate images. Moreover, a feature fusion module based on the coordinate attention (CA) idea is introduced into the teacher network, enabling it to learn features that adapt to both high-resolution and low-resolution image information, thus better guiding the training of the student network. Finally, a confidence loss function based on the gradient harmonizing mechanism (GHM) is introduced to enhance the defect detection rate. Experimental results demonstrate that the proposed ceramic substrate defect detection method based on knowledge distillation achieves an average accuracy and average recall of 96.80% and 90.01%, respectively, for the detection of five types of defect-stain, foreign matter, gold edge bulge, ceramic gap, and damage-in low-resolution (224×224) input images. Compared with current mainstream object detection algorithms, the proposed algorithm achieves better detection results.

Key words: ceramic substrate; defect detection; YOLOv5; knowledge distillation

1 引 言

陶瓷基板是大功率电力电子电路结构的基础材料。在陶瓷基板的生产过程中,常出现镀金层损伤、边缘多金、缺瓷、油污、掺杂异物五种瑕疵,这些瑕疵会影响电子器件的寿命、性能以及稳定性。快速而准确地检测并剔除这些瑕疵基板对于保证产品质量具有重要意义[1]。目前,利用机器视觉技术结合深度神经网络进行产品瑕疵检测已成为主流方法。本研究团队针对陶瓷基板瑕疵形状多变、尺寸跨度大,且多金和缺瓷瑕疵目标较小,样本量小,各类瑕疵数量分布不均匀等问题,提出了一种改进的YOLOv4算法,通过修改置信度损失函数,引入注意力机制,优化锚框尺寸等策略,实现了五类瑕疵的准确检测(平均准确率98.3%)[1]。这一精度是在单个陶瓷基板图像分辨率为608×608的条件下取得的。由于陶瓷基板尺寸较小,实际生产过程中,一个较大的基板片上需要制作多个行列排布的陶瓷基板,在保证上述图像分辨率的条件下,要完成整个基板片上所有陶瓷基板的图像采集,需要频繁移动图像采集设备,从而导致较长的图像采集时间,不利于检测速度的提高。为了提高检测速度,一个潜在的方法是扩大图像采集设备的视场(Field of View, FOV),实现多个陶瓷基板完整信息的同时采集;从而减少图像采集设备的移动频次,达到节约图像采集时间的目的。但FOV的增加,也导致了单个陶瓷基板的分辨率下降。陶瓷基板分辨率的下降(称之为低分辨率陶瓷基板图像)带来了瑕疵像素尺寸的缩小,给瑕疵(特别是小尺寸瑕疵)检测带来困难,导致大量的漏检和误检现象产生。因此,提高低分辨率陶瓷基板图像上的瑕疵检测精度对于满足实际生产中对瑕疵检测快速性的要求具有重要价值。

目前主要通过超分辨率重建方法[2]和知识蒸馏方法[3]来提高低分辨率图像上的目标检测性能。超分辨率重建方法是指利用无监督方法或者有监督方法重建出高分辨率图像,然后再利用目标检测网络对目标进行检测,该方法虽然能通过还原出高分辨率图像的特征信息从而提高检测性能,但检测高分辨率图像的速度也会随之降低,而在本文任务中,随着FOV的增加,相同时间内需要检测的图像也随之增加,因此该方法无法满足本文检测任务对实时性的要求。

知识蒸馏的基本思想是构建教师网络和学生网络,基于神经网络迁移学习的策略,利用教师网络学习到的信息指导学生网络的训练,从而使得学生网络获得更好的性能。目前知识蒸馏算法主要用于深度神经网络的模型压缩,以获得相对于教师网络更加轻量化且性能良好的学生网络[4-10]。最近,QI等[3]为提高低分辨率图像的目标检测性能,利用知识蒸馏策略将教师网络学习到的高分辨率图像特征信息用于指导学生网络的训练。取得了较好的低分辨率图像下的目

标检测性能。

受QI[3]的工作启发，本文将知识蒸馏思想引入到低分辨率陶瓷基板图像的瑕疵检测，基于YOLOv5框架构建由教师网络和学生网络组成的知识蒸馏模型YOLOv5-CSKD，利用教师网络学到的高分辨率特征信息指导学生网络的训练，从而使得学生模型能够学习到低分辨率图像特征在高分辨图像中的特征分布，同时，在教师网络中引入基于 Coordinate Attention（CA）[11]注意力思想的特征融合模块，使得教师网络学习到的特征同时适应高分辨率图像信息和低分辨率图像信息，从而能较好的指导学生网络的训练；仿真结果表明，基于知识蒸馏的陶瓷基板瑕疵检测模型 YOLOv5-CSKD 能够取得了 96.80% 的平均准确率和90.01%的平均准确率的检测性能。

2　YOLOv5算法介绍

YOLOv5算法主要由主干网络（backbone）、颈部网络（neck）、以及头部网络（head）三部分组成。其中主干网络与CSPDarknet53类似，采用多层次结构网络，提取输入图像的不同尺度的特征图，每一层均由卷积块CBS（Conv＋BatchNorm＋Silu）和残差结构组成残差块，在减少计算量，缓解梯度消失的同时保证了特征提取的完整性；颈部网络为PANet[12]，利用从底到顶的通道和从顶到底的通道，将高层和低层的特征图进行融合编码，以增强特征图的信息表达；头部网络将融合之后的特征图进行解码，得到最后的预测结果（目标的种类及位置信息）。YOLOv5的损失函数如式（1）～式（4）所示。

$$\mathrm{LOSS}=\mathrm{LOSS}_{\mathrm{conf}}+\mathrm{LOSS}_{\mathrm{cls}}+\mathrm{LOSS}_{\mathrm{reg}},\tag{1}$$

$$\mathrm{LOSS}_{\mathrm{reg}}=\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\mathrm{obj}}\left[1-\mathrm{IoU}(b_{ij},\hat{b}_{ij})+\frac{\rho^2(b_{ij},\hat{b}_{ij})}{c^2}+\alpha v\right],\tag{2}$$

$$\mathrm{LOSS}_{\mathrm{conf}}=-\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\mathrm{obj}}\left[\hat{C}_{ij}\log C_{ij}+(1-\hat{C}_{ij})\log(1-C_{ij})\right]-\lambda_{\mathrm{noobj}}\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\mathrm{noobj}}\left[\hat{C}_{ij}\log C_{ij}+(1-\hat{C}_{ij})\log(1-C_{ij})\right],\tag{3}$$

$$\mathrm{LOSS}_{\mathrm{cls}}=-\sum_{i=0}^{S^2}\sum_{j=0}^{B}1_{ij}^{\mathrm{obj}}\sum_{c\in\mathrm{classes}}\left[\hat{P}_{ij}(c)\log(P_{ij}(c))+(1-\hat{P}_{ij}(c))\log(1-P_{ij}(c))\right],\tag{4}$$

其中：LOSS 表示模型训练总的损失函数，$\mathrm{LOSS}_{\mathrm{reg}}$为定位损失函数，$\mathrm{LOSS}_{\mathrm{conf}}$为置信度损失函数，$\mathrm{LOSS}_{\mathrm{cls}}$为分类损失函数。$1_{ij}^{\mathrm{obj}}$代表预测输出中第$i$网格内第$j$预测框内有目标；$S^2$代表每个特征层上面有$S\times S$个单元格；$B=3$，代表每个单元格内有3个预测框；$b_{ij}$，$\hat{b}_{ij}$为预测框和真实框；IoU，$\rho^2$，$c^2$分别表示两个框的交并比值，中心点欧氏距离和最小闭包区域的对角线距离；v为长宽比一致性；C_{ij}，$\hat{C}_{ij}$分别为置信度的真实值和预测值；α，λ_{noobj}为权重系数；1_{ij}^{noobj}代表第i网格内第j预测框内无目标；$P_{ij}(c)$，$\hat{P}_{ij}(c)$表示目标属于每一类的概率的真实值和预测值。

最后，通过置信度阈值的设置来过滤掉得分较低的边界框，对剩下的边界框进行非极大值抑制（Non-Maximum Suppression，NMS）操作来剔除重复的边界框，从而实现目标检测。

3　基于知识蒸馏的YOLOv5模型设计

知识蒸馏是指由教师-学生网络组成的训练网络，通常是已训练好的教师模型提供知识，学生模型通过蒸馏训练来获取教师的知识。传统的目标分类或检测任务中，一般采用“0”和“1”硬标签方式对数据进行标注，但此标注方式使得标签中包含较少的类间关系，而知识蒸馏则采用教师网络的输出概率作为标签，可以较好地表达类间的相关性。

高分辨率的大尺寸图片包含着目标更详细的信息，因而可以使目标检测网络获得更好的性能。但是高分辨率图片在深度学习网络中的计算量和内存量是巨大的，实际运用时需要综合考虑计算速度与计算量。同时，在高分辨率图片上训练的网络模型无法直接用于预测低分辨率的图片。借助于知识蒸馏的方法，可以将高分辨率的复杂教师模型所学习到的知识迁移到低分辨率的高效学生模型中[13]。例如，Fu等人[14]将教师模型所学习的空间和时间知识迁移到低分辨率

的轻量级时空网络中来执行视频注意预测任务。

在利用知识蒸馏进行模型压缩时，输入图像为相同尺寸，模型无需考虑输入尺寸带来的特征尺寸不一致的问题，但在利用知识蒸馏提高低分辨率图像检测性能的方法中，由于许多网络结构采用多尺度的思想来检测不同尺寸大小的目标，使得学生网络和教师网络进行蒸馏的特征图来自于不同深度的特征层，因此需要对网络结构进行设计使得进行知识蒸馏的教师网络特征与学生网络特征的尺寸一致。

考虑到性能和实时性，本文以 YOLOv5l 作为教师网络和学生网络的主要网络结构，在此基础上实现基于知识蒸馏的目标检测算法设计。

3.1 教师网络设计

本文的知识蒸馏主要关注颈部特征层的知识蒸馏。YOLOv5 网络的输出由 3 层预测支路组成，每一层的特征尺寸和通道数都不相同，因此，当教师网络采用高分辨率图像进行训练，而学生网络采用低分辨率图像进行训练时，相同大小尺寸的特征层通道数不相同，而相同通道数的特征层则尺寸大小不同，使得无法有效进行知识蒸馏训练。为了使教师网络的特征信息能够指导学生网络的训练，需要采用卷积网络将教师网络或学生网络的特征层进行处理，使得进行知识蒸馏两个网络的特征层的尺寸和通道数在同一层级上相同。

在教师网络的训练过程中，若仅采用高分辨率图像作为训练数据，会使得教师网络学习到的特征偏向于高分辨率图像的信息分布，导致学生网络的性能提升有限，因此在教师网络的训练过程中同时利用高分辨率图像和低分辨率图像作为输入，使得教师网络不仅能学习到高分辨率图像下更精细的特征信息，也能适应低分辨率图像下的信息分布。

教师网络的结构图如图 1 所示，左侧为高分辨率图像(448 × 448)输入学习到的特征，右侧为低分辨率图像(224 × 224)输入学习到的特征，二者采用同一网络结构，其中，$C_I, C'_I(I=0,1,\cdots,5)$ 为主干网络的特征，$P_J, P'_J(J=2,3,\cdots,5)$ 为颈部网络的中间特征，$N_S, N'_S(S=2,3,\cdots,5)$ 为颈部网络的输出特征，$F_S(S=2,3,\cdots,5)$ 为融合之后的特征。由于输入图像的分辨率相差一倍，因此 N_S 的特征尺寸大小是 N'_S 的两倍，如果利用卷积对 N_S 进行尺寸缩小会使得 N_S 包含的特征信息被压缩，从而导致小目标检测能力变弱。因此，本文选择将与 N_S 特征尺寸相同

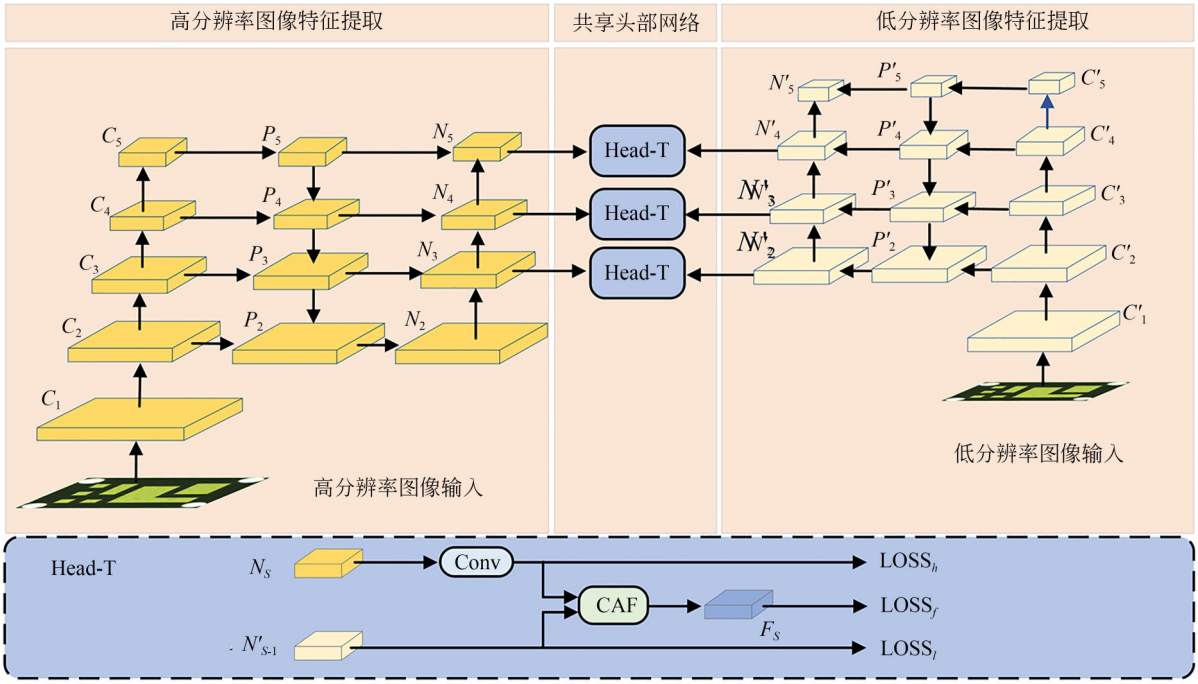


图 1 教师网络结构图

Fig. 1 Structure of teacher network

的 N'_{S-1} 进行通道压缩，使其通道数与 N_S 一致，这样就将可以来自不同分辨率的特征进行融合。

在原始 YOLOv5 模型中，当输入分辨率较小，且数据集中的目标尺寸偏小时，较少的顶层特征信息会被激活去参与预测，即 N_5 这一条支路较少参与预测，但若是直接去掉此支路则会导致性能下降，底层无法获取从 N_5 所在的层级传递的高维特征信息。基于此，考虑到 N'_5 没有对应的高分辨率图像的特征尺寸与之对应，在设计特征融合时，只利用 N'_4, N'_3, N'_2 参与特征融合以及预测。由于 N'_2 对应的特征尺寸更大，有着更加精细化的特征，能在一定程度上提高小目标检测的能力。

教师网络的损失函数如式(5)所示，其中 $\mathrm{LOSS}_h, \mathrm{LOSS}_l, \mathrm{LOSS}_f$ 分别为高分辨率图像的特征、低分辨率图像的特征以及两种分辨率图像的融合特征经过 YOLOv5 头部网络所产生的预测损失(通过公式(1)计算得到)；λ 为超参数，用来平衡损失中融合特征损失所占的比重。

$$\mathrm{LOSS}_T = \mathrm{LOSS}_h + \mathrm{LOSS}_l + \lambda \times \mathrm{LOSS}_f. \quad (5)$$

3.2　特征融合模块设计

为了使指导学生网络训练的教师网络特征同时能包含两种分辨率图像的信息，QI[3]设计了一个特征融合模块，将来自不同分辨率图像的特征进行融合生成新的特征并参与教师网络的损失函数计算，最后用这个融合特征指导学生网络的训练，但 QI 的特征融合方法仅给来自不同分辨率的特征赋予一个权重，无法从多维数据中选择出更加全面有效的特征信息，因此本文基于 CA 注意力的思想，提出了一个能够同时关注通道和空间权重的轻量级特征融合模块，即 CAF (Coordinate Attention Fuse)模块，模块结构图如图 2 所示，其中，$N_S, N_{S-1}', F_S (S=3,4,5)$ 分别为教师网络、学生网络的颈部结构输出特征和二者的融合特征。

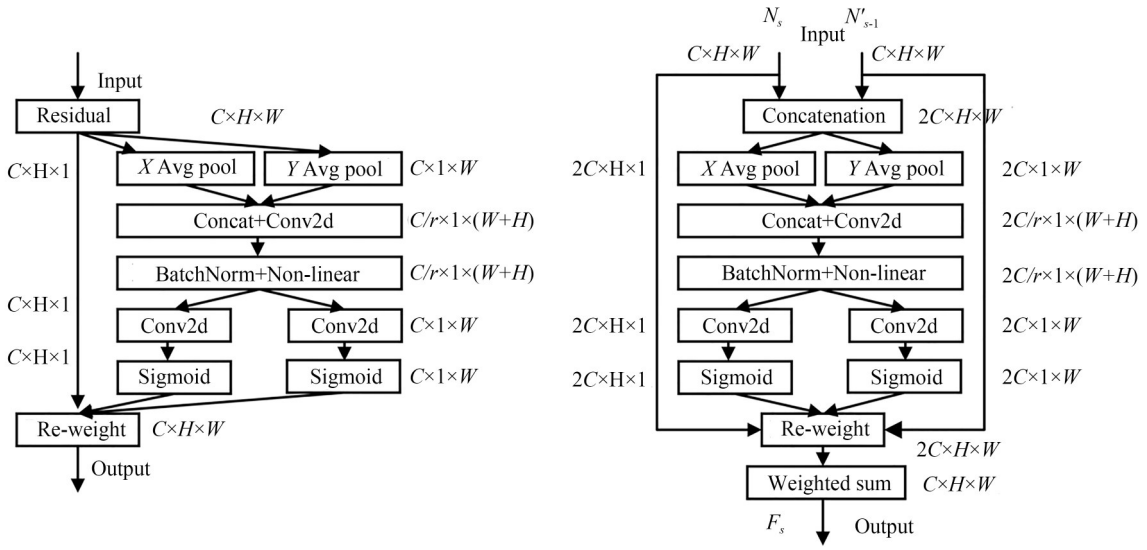


图 2　特征融合模块结构图

Fig. 2　Structure of feature fusion module

假设输入为 N_S, N_{S-1}，其中 $\{N_S, N_{S-1}\} \in R^{C\times H\times W}$，$C$ 为特征通道数，H, W 为特征尺寸的高和宽。首先，将 N_S, N_{S-1} 进行通道拼接得到 $N_S^{\mathrm{con1}} \in R^{2C\times H\times W}$，然后分别沿着 X，Y 方向进行 Avg Pool 得到 $N_S^x \in R^{2C\times H\times 1}$，$N_S^y \in R^{2C\times 1\times W}$，$N_S^x, N_S^y$ 分别反映了特征层在 X, Y 方向上的关联性；将 N_S^x, N_S^y 处理成同样的维度并进行通道拼接，然后经过卷积操作，Non-linear 层之后，得到 $N_S^{\mathrm{con2}} \in R^{2C/r\times 1\times (H+W)}$，使得特征层通道信息和空间信息发生了交互；将 N_S^{con2} 进行分离并经过卷积层以及 Sigmoid 函数之后，得到 $N_S^{x2} \in R^{2C\times H\times 1}$，$N_S^{y2} \in R^{2C\times 1\times W}$，此时的 N_S^{x2}, N_S^{y2} 融合了两个特征之间的空间以及通道信息，然后将二者相乘得到权重矩阵 $h_S \in R^{2C\times H\times W}$，分别将 N_S, N_{S-1} 与 $h_S\{0\}, h_S\{1\}$ 进行矩阵相乘，然后将得到的结果相加得到最终的融合特征 F_S，其中，$h_S\{0\}, h_S\{1\}$ 表示按 h_S 的通道维度上的索引顺序将 h_S 分成维度为 $C\times H\times W$ 的两个特征块。

最后得到的融合特征 F_S 不仅从高分辨率特征和低分辨率特征中选取了最合适的特征信息，而且同时考虑了通道和空间上的最优特征信息。

3.3 学生网络设计

学生网络的结构图如图 3 所示，学生网络主要利用低分辨率图像进行检测，其主体网络结构与教师网络一致，无特征融合模块，在训练时直接加载教师网络的权重进行训练。

学生网络的损失函数如式(6)所示，学生网络的损失主要包含两部分，一部分是学生网络自身低分辨率图像输入下的原始 YOLOv5 损失，即 LOSS_l；另一部分是学生网络颈部网络输出的特征 N_{S-1} 与教师网络的融合特征 F_S 之间的知识蒸馏损失，即 $\mathrm{LOSS}_{\mathrm{KD}}$，其定义如式(7)所示，其中 $S\in[3,5]$，$\mathrm{LOSS}_{\mathrm{KD}}$ 采用 L1 损失度量 N_{S-1} 与 F_S 之间的差距；γ 为超参数，用来平衡学生网络的损失函数中预测损失与知识蒸馏损失所占的比重。

$$\mathrm{LOSS}_S=(1-\gamma)\times\mathrm{LOSS}_l+\gamma\times\mathrm{LOSS}_{\mathrm{KD}},\quad(6)$$

$$\mathrm{LOSS}_{\mathrm{KD}}=\sum_S\left|F_S-N'_{S-1}\right|,S\in[3,5].\quad(7)$$

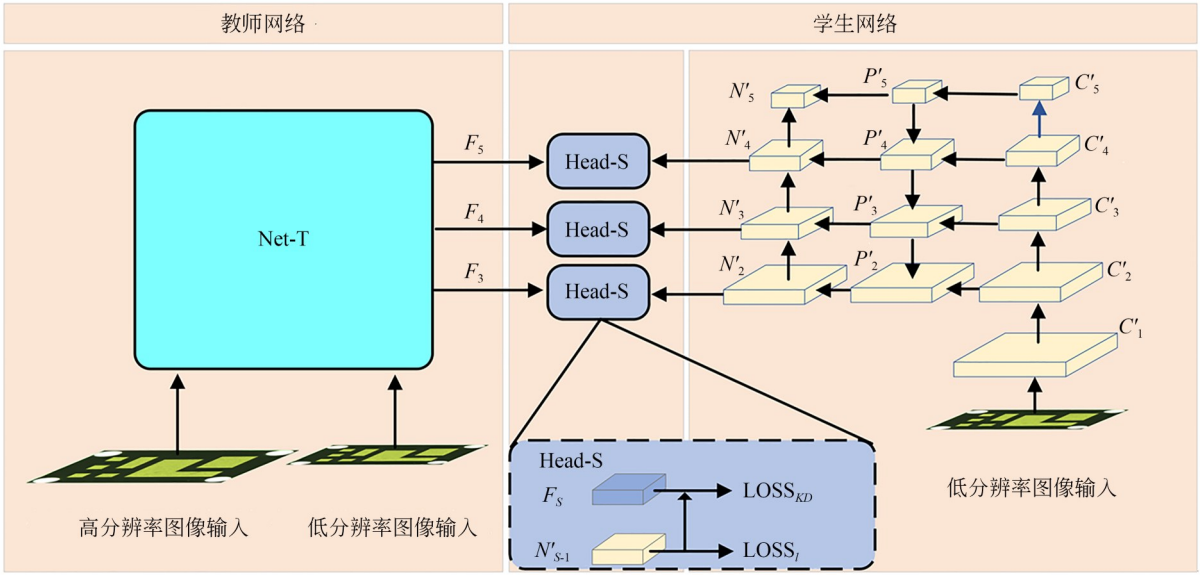


图 3 学生网络结构图

Fig. 3 Structure of student network

3.4 基于 GHM 机制的置信度损失函数设计

当模型训练至稳定时，一些类内差距较大或者数量较少的难检测瑕疵样本，其损失会被其他同类的易检测样本或数量较多的其他类样本所抑制，使得其特征在训练过程中难以被有效学习，从而导致模型对这些瑕疵的特征信息不敏感，造成漏检。为了缓解这种情况，本文利用 GHM 分类损失函数(gradient harmonizing mechanism-classification, GHM-C)[15]改进了模型的置信度损失函数，学生网络和教师网络的置信度损失函数均采用改函数。具体实现步骤为：

(1)对于模型输出的第 k 个预测框($k=1,\cdots N$，N 为预测框的总数)，按公式(8)计算其梯度模 g_k，其中 C_k，$\hat{C}_k$ 分别代表第 k 个预测框的置信度预测值与真实值；

(2)将置信度的输出区间[0,1]按步长划分为若干等长子区间 $S_t(t=1,\cdots M$，M 为子区间数量)，统计预测框的梯度模落在子区间 S_t 的数量 N_t；计算得到各样本的梯度密度调和参数 β_k(如式(9)所示)；

(3)将原始模型计算的置信度交叉熵的梯度值分别乘上各自归属子区间的梯度密度调和参数 β_k，从而达到衰减其影响的目的(如式(10)所示)。

$$g_k=\left|C_k-\hat{C}_k\right|=\begin{cases}1-C_k, & \text{if } \hat{C}_k=1\\ C_k, & \text{if } \hat{C}_k=0\end{cases},\quad(8)$$

$$\beta_k=\frac{N}{N_t}\varepsilon,\ \text{if } g_k\in S_t,\quad(9)$$

$$\mathrm{L}_{\mathrm{GHM-C}}=\frac{1}{N}\sum_{k=1}^{N}\beta_k\mathrm{L}_{\mathrm{CE}}(C_k,\hat{C}_k),\quad(10)$$

$$LOSS = LOSS_{reg} + L_{GHM\text{-}C} + LOSS_{cls}, \quad (11)$$

其中：$L_{GHM\text{-}C}$ 表示最终的基于 GHM-C 的置信度损失函数，L_{CE} 表示交叉熵损失。修改之后模型最终的损失函数如式(11)所示，LOSS 是总的损失函数，其中 $LOSS_{reg}$ 为定位损失函数，$LOSS_{cls}$ 为分类损失函数。最终的教师网络和学生网络的损失函数 $LOSS_h$，$LOSS_l$，$LOSS_f$ 均由 LOSS 组成。

4　实验及结果分析

4.1　实验环境

本次实验的操作系统为 Windows 10(64 位)，CPU 为英特尔 Core i9-10900X@3.70GHz，内存为 64 GB，GPU 为两个 NVIDIA GeForce RTX 3090，总显存为 48 GB。深度学习框架实现为 Python Pytorch，GPU 加速库为 CUDA11.0。

4.2　数据集介绍

本文需要检测的陶瓷基板图像如图 4(a)所示。其中方块区域为镀金层，本文所指的瑕疵都是指出现在镀金层上的瑕疵。典型瑕疵主要有污渍、异物、多金、缺瓷以及损伤五类，部分瑕疵如图 4(b)所示。污渍瑕疵会出现镀金层上任何一个区域，由于无法预测是何种物质导致的污渍，因此无法定义一个通用的颜色信息或形状信息去描述这种瑕疵，且污渍的尺寸跨度大、形状多变；异物瑕疵大多来源于外界环境，因此形状和颜色也具有多样性及不定性，且由于陶瓷基板属于高集成度的产品，图像采集均是采用带有显微功能的相机完成，因此一些极细小的物质也会被认为是异物，并且很多异物的尺寸很小，给模型检测带来了很大的挑战；缺瓷瑕疵主要分布在镀金层边缘，主要是由于生产工艺不良造成镀金不全，其形状较小，属于小目标瑕疵；损伤瑕疵主要是镀金层的划伤或者刮伤，尺寸跨度大且形状(线状或带状)分布多变；多金瑕疵主要是由于设备或生产工艺问题造成的镀金层边缘出现凸起，形状较小，也属于小目标瑕疵，且分布区域多变。

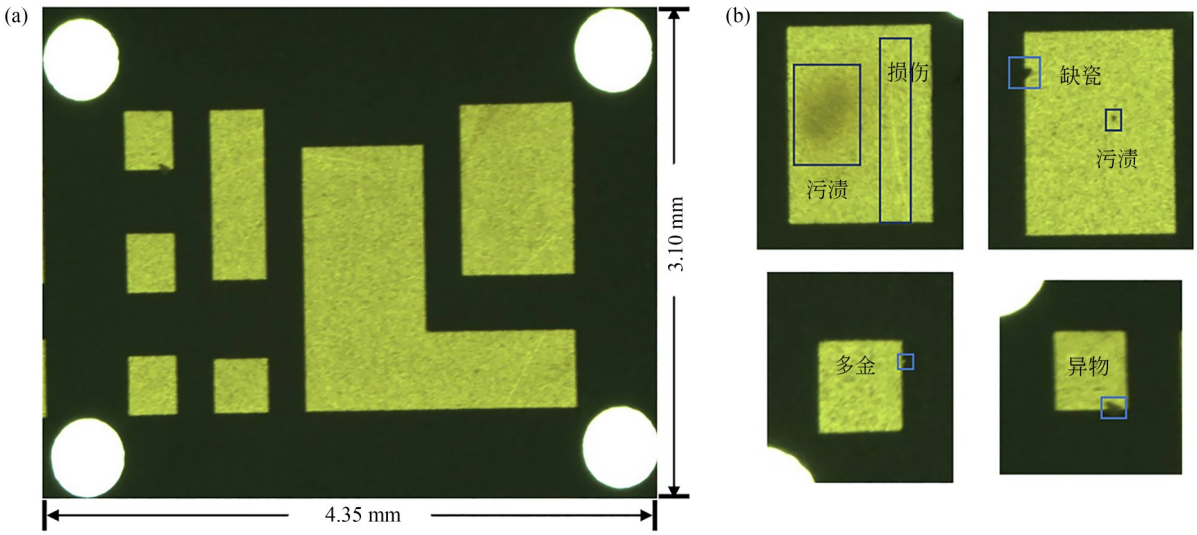


图 4　陶瓷基板与部分瑕疵示意图

Fig. 4　Schematic diagram of ceramic substrate and defects

采集的单张陶瓷基板原始图像经过裁剪后分辨率约为 1 637×1 175，按照生产需求，采集的图像由原来的一次性采集 1 个陶瓷基板到现在的一次性采集 16 个陶瓷基板，如图 5 所示。裁剪后的单张陶瓷基板原始图像分辨率将降至 410×294 左右，因此模型输入分辨率需降至 224×224 左右。为了还原真实现场的低分辨率陶瓷基板瑕疵检测，本文实验的数据集由裁剪后的原始图片分别降采样至 224×224 分辨率与 448×448 分辨率，其中 224×224 为低分辨率输入图像，448×448 为高分辨率输入图像。

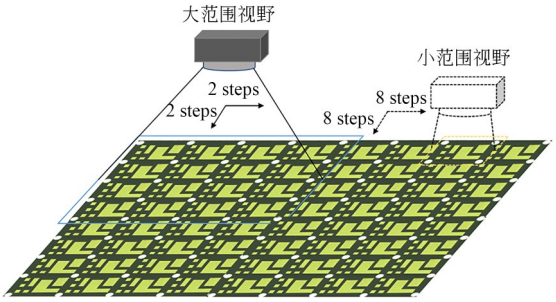


图5 相机拍照视野示意图

Fig. 5 Schematic diagram of camera field of view

4.3 数据集增强

本实验共采集到2 413张陶瓷基板图像，经过目标区域提取预处理之后得到陶瓷基板数据集。按照文献[16]的方式对2 215张训练样本进行数据增强，具体步骤为：从2 215张图片中裁剪出50张包含典型瑕疵（多为样本数较少的瑕疵）的子区域（每个区域只包含一种典型缺陷）；并将这50张子区域图像进行水平、垂直翻转，尺度的缩放，引入高斯噪声（标准差为1.2）以及锐化操作等；最后将这些增强后的裁剪图像“粘贴”至无瑕疵陶瓷基板图像中可能出现瑕疵的区域，从而获得增强后的图片2 386张，用作训练，剩余198张未增强图片用作测试。使用LabelImg图像标注软件进行人工标注，数据集格式为VOC2007，后转为YOLO数据集格式。

4.4 网络训练

本文的网络训练分为教师网络训练和学生网络训练。教师网络训练阶段，总训练Epoch为200次，批量大小（Batch size）为32，初始学习率为0.01，warm_up的Epoch数为3、动量为0.8、初始bias_lr为0.1，权重衰减系数为0.000 5，使用随机梯度下降法（SGD）进行优化，动量等于0.937。本文将超参数λ设为0.8。

学生网络训练阶段，总训练Epoch为200次，批量大小（Batch size）为64，初始学习率为0.01，warm_up的Epoch数为3、动量为0.8、初始bias_lr为0.1，权重衰减系数为0.0005，使用随机梯度下降法（SGD）进行优化，动量等于0.937，知识蒸馏采用的温度系数为3.0。本文将超参数γ设为0.4。

4.5 实验结果

4.5.1 不同检测算法的结果对比

本文选取准确率、召回率作为评判模型性能的指标，计算公式如式（12），式（13）所示：

$$\text{Precision}=\frac{\text{TP}}{\text{TP}+\text{FP}},\tag{12}$$

$$\text{Recall}=\frac{\text{TP}}{\text{TP}+\text{FN}},\tag{13}$$

其中：TP表示将正例预测为正例的个数，FP表示将反例预测为正例的个数，FN表示将正例预测为反例的个数。

待检测图片数量共198张，每张图片包含1至5种类别的瑕疵，污渍、异物、多金、损伤以及缺瓷五种瑕疵的检测结果如表1所示。

由表1可知，本文提出的算法对于陶瓷基板中污渍、异物、多金、损伤以及缺瓷五种不同种类瑕疵的检测准确率已基本达到87.88%以上，召回率均已基本达到85.29%以上。其中，多金和缺瓷的检测效果较差，其原因可能为：一方面这两种瑕疵的样本数量较少，因此在训练时势必会受到影响；另一方面，这两种瑕疵尺寸相对较小，当图片分辨率降低时，模型能够关注的像素变少，因此检测性能较差。

表1 瑕疵检测结果

Tab. 1 Results of defect detection

瑕疵类别	总数	正确检测	误检	漏检	Precision /%	Recall /%
污渍	1 591	1 441	33	150	97.76	90.57
异物	28	25	2	3	92.59	89.28
多金	34	29	4	5	87.88	85.29
损伤	310	276	15	34	94.84	89.03
缺瓷	120	104	8	16	92.86	86.67
合计	2 083	1 875	62	208	96.80	90.01

为了进一步验证本文算法在陶瓷基板瑕疵检测上的有效性，将本文算法与常见的目标检测算法Faster R-CNN[17]，Cascade RCNN[18]，YOLOv4[19]，YOLOX[20]，YOLOv6[21]，YOLOv7[22]以及改进前的YOLOv5在本文数据集上进行对比试验，采用原文作者的实验参数，未采用知识蒸馏的模型均训练400 Epoch，选取平均准确率、平均召回率作为评价指标，不同算法的性能如表2所示。表2中的“-”表示该模型在给定的训练次数下实验未收敛。

根据表2，本为提出的YOLOv5-CSKD算法

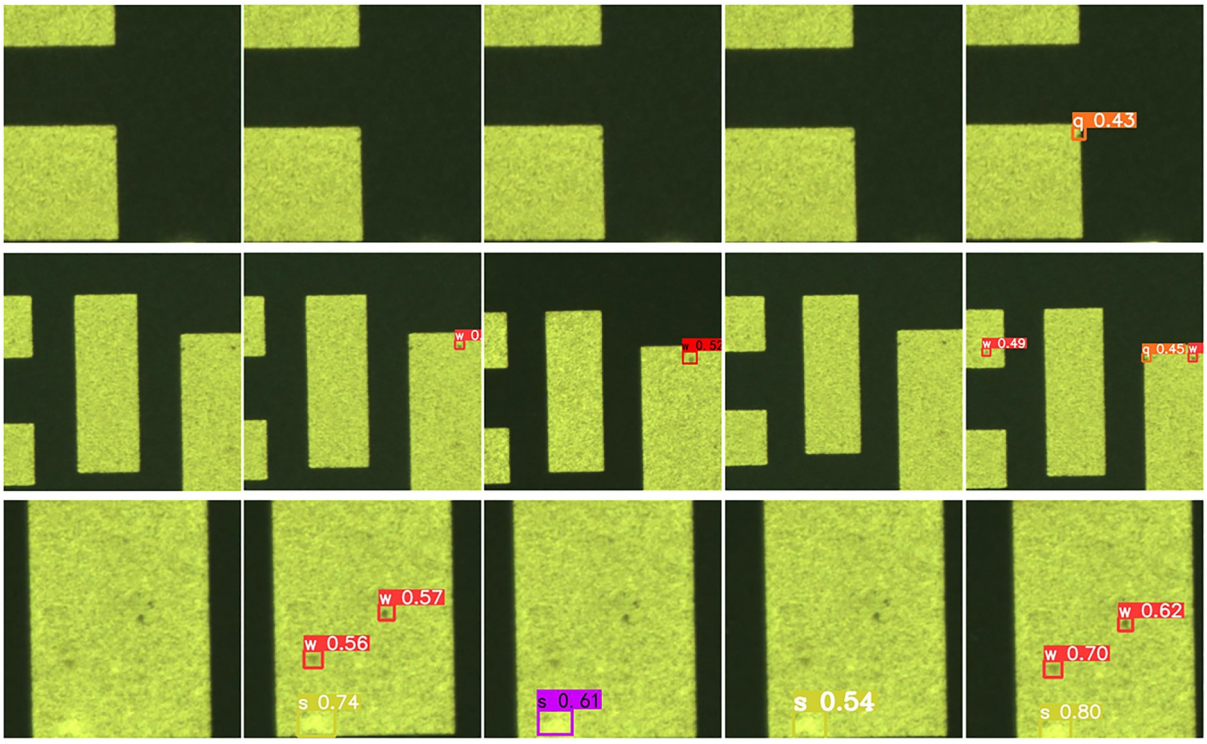
对陶瓷基板瑕疵检测平均准确率和平均准确率分别达到了 96.80% 和 90.01%，相比于原始的 YOLOv5 算法，分别提高了 6.99% 和 8.81%，且相比于原始 YOLOv5 模型，本文提出的 YOLOv5-CSKD 算法在 224×224 分辨率下的检测时间相差较小。在 224×224 分辨率下所有对比算法中，YOLOv5-CSKD 具有最好的性能，说明本文提出创新的有效性。

表 2　不同算法瑕疵检测结果

Tab. 2　Defect detection results of different algorithms

模型	输入尺寸	Precision/%	Recall/%	检测速度/(ms·f^{-1})
Faster R-CNN	(224×224)	—		—
Cascade RCNN	(224×224)	—		—
YOLOv4	(224×224)	85.86	75.89	126.59
YOLOv5	(224×224)	89.12	81.20	83.02
YOLOv5	(448×448)	89.81	84.33	93.45
YOLOX	(224×224)	85.89	77.79	115.72
YOLOv6	(224×224)	88.44	79.35	87.42
YOLOv7	(224×224)	—	—	—
MSAD	(224×224)	—	—	—
YOLOv4-CS	(224×224)	93.12	85.43	142.70
YOLOv5-CSKD	(224×224)	96.80	90.01	87.04

图 6 给出了陶瓷基板典型瑕疵输入为 224×224 分辨率的检测结果(局部截图)。图 6 中(a)～(e)的检测结果依次为 YOLOv4，YOLOv5，YOLOX，YOLOv6，YOLOv5-CKSD。



(a) YOLOv4　(b) YOLOv5　(c) YOLOX　(d) YOLOv6　(e) YOLOv5-CSKD

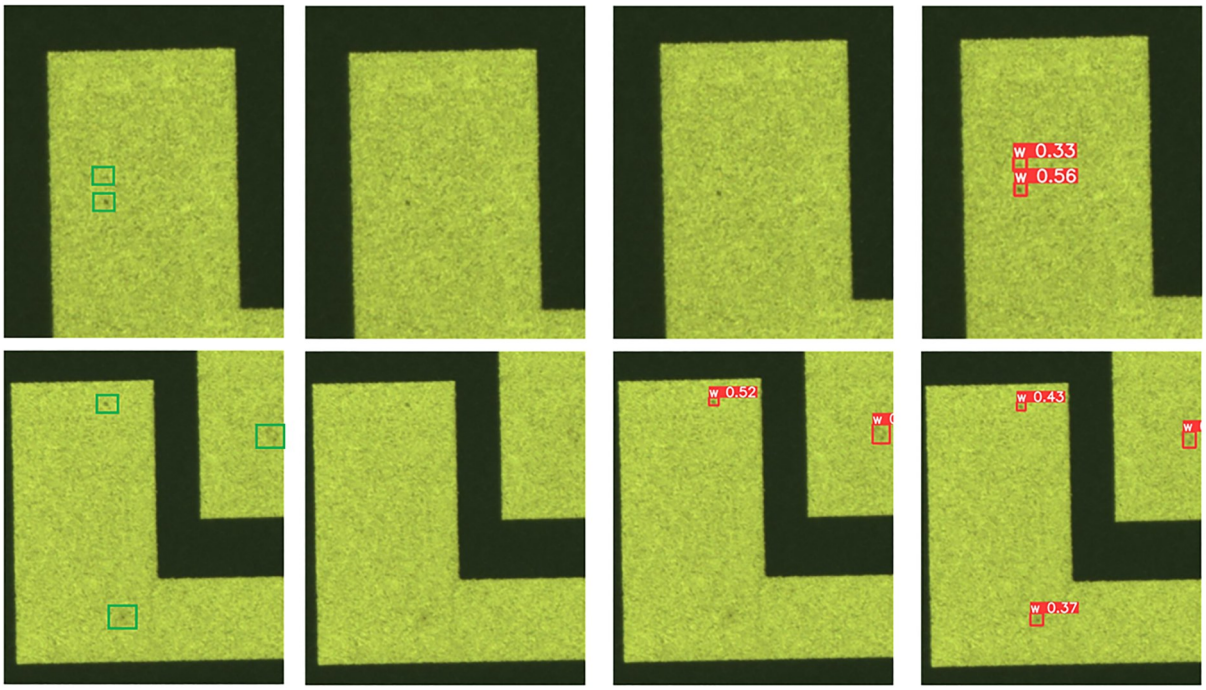
图 6　不同模型的局部检测结果

Fig. 6　Local detection results of different models

从图 6 中可以看出，在(224 × 224)的输入分辨率下，YOLOv4 的检测结果相对较差。原始 YOLOv5，YOLOX，YOLOv6 均存在不同程度的漏检，其中，原始的 YOLOv5 检测结果明显好于 YOLOX，YOLOv6，所以本文采用 YOLOv5 作为基准研究模型。对比原始 YOLOv5 以及 YOLOv5-CKSD，可以得出，YOLOv5-CKSD 能有效检测出原始 YOLOv5 模型无法检测出的瑕疵，因此可以证明本文提出的模型 YOLOv5-CKSD 在低分辨率陶瓷基板输入下的检测性能较其他方法有明显提升。

4.5.2　对比实验

为了验证本文提出的特征融合模块 CFA 的有效性，本文针对该模块进行了对比实验。对比实验结果如表 3 所示，其中，SUM 表示低分辨率特征和高分辨特征直接逐像素相加，C-FF 为 QI[3] 提出的特征融合方式，CAF 为本文提出的特征融合方式。三种融合特征融合方式的局部对比结果如图 7 所示。



(a) 真值
(a) Ground truth

(b) SUM

(c) C-FF

(d) CAF

图 7　不同特征融合方式的局部检测结果

Fig. 7　Local detection results of different feature fusion methods

从表 3 中可以看出，在对 YOLOv5 进行知识蒸馏时，基于 SUM 融合方式与基于 C-FF 融合模块的性能差距较小，本文提出 CAF 模块具有更好的结果，获得了 96.80% 的平均准确率和 90.01% 的平均准确率。从图 7 中可以看出，在低分辨率输入下，SUM 和 C-FF 对小目标信息的敏感度较低，无法准确检测出瑕疵，而本文提出的 CAF 模块则能准确检测出图像中的瑕疵。

表 3　不同特征融合方式的实验结果

Tab. 3　Experimental results of different feature fusion methods

融合方式	Size	Precision/%	Recall/%
原始 YOLOv5	224 × 224	89.12	81.20
SUM	224 × 224	95.01	88.34
C-FF	224 × 224	95.13	88.14
CAF	224 × 224	96.80	90.01

4.5.3　消融实验

为了验证在置信度损失函数中添加GHM-C对模型性能提升的有效性，本文分别进行了消融实验，实验采用的模型使用不同置信度损失函数，其他设置以及参数相同，模型输入分辨率为224×224，消融实验结果如表4所示。

表4　采用不同置信度损失函数的对比实验结果

Tab.4　Comparative experimental results using different confidence loss functions

教师网络采用GHM-C	学生网络采用GHM-C	平均准确率/%	平均召回率/%
		89.79	84.01
√		86.53	82.33
	√	86.22	88.35
√	√	96.80	90.01

从表中可以看出，只给教师网络采用GHM-C会导致学生网络的平均每准确率和平均召回率略有降低，可能原因为教师网络和学生网络采用的置信度损失函数不同，因此使得知识蒸馏效果产生负面效果；只给学生网络添加GHM-C使得学生网络学习的特征与教师网络有所不同，从而导致学生网络无法从教师网络获得正确的信息，因此学生网络的平均召回率明显提升，但是预测出许多错误的结果，平均准确率略有降低；同时给教师网络和学生网络使用GHM-C不仅使得模型的召回率提升，同时准确率也有提升，在之前组合中许多被误检的目标在这个组合中能够准确地被检测出。因此本文提出的YOLOv5-CSKD在教师网络中和学生网络中均采用GHM-C作为置信度损失函数训练。

5　结　论

针对目前陶瓷基板瑕疵检测在低分辨率输入图像上不足之处，论文将知识蒸馏引入到陶瓷基板瑕疵检测中，基于YOLOv5网络模型设计了教师网络以及学生网络，并设计了特征融合模块同时融合高分率图像特征和低分辨率图像特征，用融合之后的特征对学生网络进行知识蒸馏，可以使得学生网络学习到更加全面的信息。实验结果表明，基于知识蒸馏的陶瓷基板瑕疵检测算法对低分辨率陶瓷基板图像的五种不同类型的瑕疵均取得更好地检测效果，平均准确率和平均准确率分别达到了96.80%和90.01%。本文仅较多考虑了YOLOv5网络与知识蒸馏算法的结合，在后续的研究中，我们将对网络结构进行优化改进，实现检测性能更优、模型更轻量化。

参考文献：

[1] 郭峰，朱启兵，黄敏，等．基于改进YOLOV4的陶瓷基板瑕疵检测[J]．光学 精密工程，2022，30(13)：1631-1641.
GUO F，ZHU Q B，HUANG M，*et al*. Defect detection in ceramic substrate based on improved YOLOV4[J]. *Opt. Precision Eng.*，2022，30(13)：1631-1641.(in Chinese)

[2] LI J C，FANG F M，MEI K F，*et al*. *Multi-Scale Residual Network for Image Super-Resolution*[M]. Computer Vision - ECCV 2018. Cham：Springer International Publishing，2018：527-542.

[3] QI L，KUEN J，GU J X，*et al*. Multi-scale aligned distillation for low-resolution detection[C]. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 20-25，2021，*Nashville*，*TN*，*USA*. IEEE，2021：14438-14448.

[4] HINTON G，VINYALS O，DEAN J. Distilling The Knowledge in a Neural Network[EB/OL]. 2015：*arXiv*：1503.02531. https://arxiv.org/abs/1503.02531.pdf

[5] HE T，SHEN C H，TIAN Z，*et al*. Knowledge adaptation for efficient semantic segmentation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 15-20，2019，*Long Beach*，*CA*，*USA*. IEEE，2020：578-587.

[6] LIU Y F，SHU C Y，WANG J D，*et al*. Structured knowledge distillation for dense prediction[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*，2023，45(6)：7035-7049.

[7] CHEN G B，CHOI W，YU X，*et al*. Learning efficient object detection models with knowledge distillation[C]. *Proceedings of the* 31*st International Conference on Neural Information Processing Systems. December* 4-9，2017，*Long Beach*，*California*，*USA*. New York：ACM，2017：742-751.

[8] WANG T，YUAN L，ZHANG X P，*et al*. Distill-

ing object detectors with fine-grained feature imitation[C]. 2019 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*). 15-20, 2019, *Long Beach*, *CA*, *USA*. IEEE, 2020: 4928-4937.

[9] ZHANG L, MA K. Improve object detection with feature-based knowledge distillation: towards accurate and efficient detectors[C]. *International Conference on Learning Representations*, *Addis Ababa*, *Ethiopia*. 2020.

[10] 褚晶辉, 史李栋, 井佩光, 等. 适用于目标检测的上下文感知知识蒸馏网络[J]. 浙江大学学报(工学版), 2022, 56(3): 503-509.
CHU J H, SHI L D, JING P G, *et al*. Context-aware knowledge distillation network for object detection[J]. *Journal of Zhejiang University* (*Engineering Science*), 2022, 56(3): 503-509. (in Chinese)

[11] 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*). 20-25, 2021, *Nashville*, *TN*, *USA*. IEEE, 2021: 13708-13717.

[12] 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*. 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 8759-8768.

[13] 黄震华, 杨顺志, 林威, 等. 知识蒸馏研究综述[J]. 计算机学报, 2022, 45(3)624-653.
HUANG Z H, YANG S Z, LIN W, *et al*. Knowledge distillation: a survey[J]. *Chinese Journal of Computers*, 2022, 45(3)624-653. (in Chinese)

[14] FU K, SHI P P, SONG Y F, *et al*. Ultrafast video attention prediction with coupled knowledge distillation[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, 34(7): 10802-10809.

[15] LI B Y, LIU Y, WANG X G. Gradient harmonized single-stage detector[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, 33(1): 8577-8584.

[16] KISANTAL M, WOJNA Z, MURAWSKI J, *et al*. Augmentation for Small Object Detection[EB/OL]. 2019: *arXiv*: 1902.07296. https://arxiv.org/abs/1902.07296.pdf

[17] 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.

[18] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection [C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 18-23, 2018, *Salt Lake City*, *UT*, *USA*. IEEE, 2018: 6154-6162.

[19] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. 2020: *arXiv*: 2004.10934. https://arxiv.org/abs/2004.10934.pdf

[20] GE Z, LIU S, WANG F, *et al*. YOLOX: Exceeding YOLO Series in 2021[EB/OL]. 2021: *arXiv*: 2107.08430. https://arxiv.org/abs/2107.08430.pdf

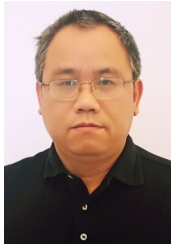
[21] LI C, LI L, JIANG H, *et al*. YOLOv6: A Single-Stage Object Detection Framework for Industrial Applications [EB/OL]. 2022: *arXiv*: 2209.02976. https://arxiv.org/abs/2209.02976.pdf

[22] 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 [EB/OL]. 2022: *arXiv*: 2207.02696. https://arxiv.org/abs/2207.02696.pdf

作者简介:

郭 峰(1998—),男,湖北麻城人,硕士研究生,2020年武汉科技大学获得学士学位,主要从事深度学习目标检测方面的研究。
E-mail: guofeng20_20@163.com

通讯作者:



朱启兵(1973—),男,安徽合肥人,教授,博士生导师,2006年于东北大学获得博士学位,主要从事传感与检测技术、信息感知与智能处理、物联网技术等领域的研究工作。
E-mail: zhuqib@163.com