DOI:10.19652/j.cnki.femt.2305406

基于深度学习的密集物料检测方法*

朱 希 李 燕 施林枫
(南京信息工程大学自动化学院 南京 210044)

摘 要:工业密集物料检测对计数精度有着较高的要求。对于检测样本疏密程度较均匀且互相无遮盖的情况,传统方法检测效果尚好,但是针对疏密分布不均匀或者样本互相之间存在遮盖的场景误识别的情况便较为严重。因此,为了提高检测的准确率与计数的精度,在 HRNet 的基础上进行改进,提出一种自注意力多尺度融合模型,主模型使用 HRNet 在不同分辨率特征图之间进行特征交互融合,同时在高分辨率特征图上添加自注意力机制加强模块增强全局特征提取。其次,针对物料中包含少数大料但密集检测检测效果较差的情况,采用了双通道物料大小判断执行机制,添加了 YOLO 框架对物料大小分类进行检测。最后,数据集均由 X 射线无损检测设备进行采集标注,在此数据集上模型的预测精度达到了 96.7%,相较于其他模型有较大的提升。

关键词:密集检测;自注意力机制;双通道判断机制;X 射线无损检测
中图分类号: TN2 **文献标识码**:A **国家标准学科分类代码**: 520.604

Dense material detection method based on deep learning

Zhu Xi Li Yan Shi Linfeng
(Automation College, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Industrial intensive material inspection has high requirements for counting accuracy. Traditional methods perform well in cases where the density of detected samples is relatively uniform and they do not overlap. However, in scenarios with uneven density distribution or overlapping samples, the issue of misrecognition becomes more severe. To enhance detection accuracy and counting precision, this study improves upon the HRNet architecture and proposes a self-attention multi-scale fusion model. The main model employs HRNet for feature interaction fusion among different resolution feature maps and enhances global feature extraction by adding a self-attention mechanism to high-resolution feature maps. Furthermore, to address situations where the inspection performance is poor for materials with few but large components, a dual-channel material size determination mechanism is introduced, utilizing the YOLO framework for material size classification detection. Lastly, the datasets used in this study are collected and annotated using X-ray non-destructive testing equipment. The proposed model achieves a prediction accuracy of 96.7% on this dataset, demonstrating improvement compared to other models.

Keywords: dense inspection; self-attention mechanism; dual-channel decision mechanism; X-ray non-destructive testing

0 引 言

近年来,随着半导体行业的不断发展。半导体物料的精密化和微小化趋势不断加强。传统的检测计数方法针对固定尺寸且分布均匀的物料表现良好。然而,在处理相对高分辨率图像进行检测和计数时,传统方法的效果较差。因为精密物料通常具有非常小的尺寸并且其分布差异很大,传统算法难以在这种场景进行精确计数。目前,基于深度学习的计数方法[1]已经逐渐取代基于传统的技术方法成为密集检测计数方向的主流算法。

深度学习方面的密集检测算法主要分成单阶段算法[2]和双阶段算法[3]两类。单阶段模型的工作流程是先

收稿日期:2023-08-07
* **基金项目**:南京信息工程大学创新创业基金(WXCX202125)、江苏省研究生科研与实践创新计划项目(SJCX23_0369)资助

从图像中生成预测候选框，再通过算法对候选框进行条件筛选，最终得到预测框。基本检测方法包括 RCNN[4]、Fast-RCNN[5]、Mask-RCNN[6]、FPN[7]等。双阶段模型能够提供更精确的检测结果，但这种精确性通常伴随着更复杂的运算过程和较长的运行时间。与之相比，单阶段模型省略了生成候选框的步骤，而是直接对图像中的特征点进行检测和定位，因此通常具有更快的运算速度，从而提高了检测效率。2016 年 Shang 等[8]使用 GoogleNet 作为主干网络对图像进行特征提取在进行解码操作，可以实现对特定区域的计数需求。2017 年 Zhang 等[9]提出通过生成密度图使用端到端的模式对模型进行训练，在人群计数及车辆计数方面效果得到改善。近几年注意力机制[10]在计算机视觉领域得到广泛应用，Hossain 等[11]使用注意力机制去提升模型的工作效率，通过全局注意力与局部注意力分别去关注图像内与图像之间的特征，减少了噪声对检测结果的影响，提升了检测精度。由此可见基于传统机器学习的计数方法正逐步向基于深度学习的方法转变，并在计数的准确性上取得了实质性的进展[12]。但是深度学习方法也存在局限性，在面对样本之间间隔较大较为稀疏的场景进行识别时检测较为精准，在处理较为密集的场景或是分布不均匀的情况时，密度图是由一系列模糊的高斯斑点构成且存在样本重叠的现象，现有的深度学习方法很难做到精准定位与计数。

针对上述问题，本文以 HRNet[13]网络模型为基础进行改进。在骨干网络上添加自注意力模块[14]以增强全局特征提取的能力。此外，还引入了双通道物料大小判断执行机制，以确保少量大块物料计数的准确性。这些改进使得在稍微增加计算时间的情况下，显著提高了精确度。

1 本文方法

本文以 HRNet 为基础设计了基于自注意力机制的双通道并行检测模型，网络结构如图 1 所示。该网络结构由主干网络与并行网络组成。主干网络使用 HRNet 进行特征提取，并行网络使用基于 Transformer 的高分辨率特征模块(high resolution interaction transformer，HRIT)同时加强全局特征提取，利用聚焦反距离变换算法[15](focus inverse distance transform，FIDT)对特征图进行映射得到 FIDT 变换图(focus inverse distance transform map，FIDTM)，再通过局部极大值检测方法求出样本数量并确定目标位置。针对电子物密集且分布均匀的情况进行了两项改进：1)提出了基于自注意力机制的多尺度融合模型 HRIT-HRNet。在网络原有高分辨率的执行通道上添加了并行的 HRIT block[16]对特征进行加强提取。确保了图片在生成高分辨率图形特征时的准确性与快速性。2)采用了双通道物料大小判断执行机制，考虑到物料存在体积较大的情况，密集检测会将单个大块物料判断为多个小物料，影响计数精度。额外添加 YOLO v5[17-18]算法对输入物料进行大小区分并且对大块物料进行检测，提高算法精度。

1.1 双通道自注意力多尺度融合模型

HRIT-HRNet 的设计需求是牺牲少量时间对图像特征进行更全面的提取。对此本文选择了 HRNet 作为基础模型。该模型分成 4 个 stage，每个 stage 通过交叉尺度融合对输入图片进行逐层特征提取，以不同分辨率融合输出，为了实现轻量化同时又保证特征完整提取，本文算法在每个 stage 节点之间加入高分辨率特征交互模块 HRIT，将提取的注意力权重与下方网络特征融合输出进行聚合输出。

该网络主干分成了 4 个阶段。第 1 个阶段是构建包含不同尺度的特征图，这些特征图会输入到不同分辨率的卷积分支中，卷积分支负责对接收到的特征图进行卷积运算，以生成不同分辨率的特征图，以供后续的信息交互和

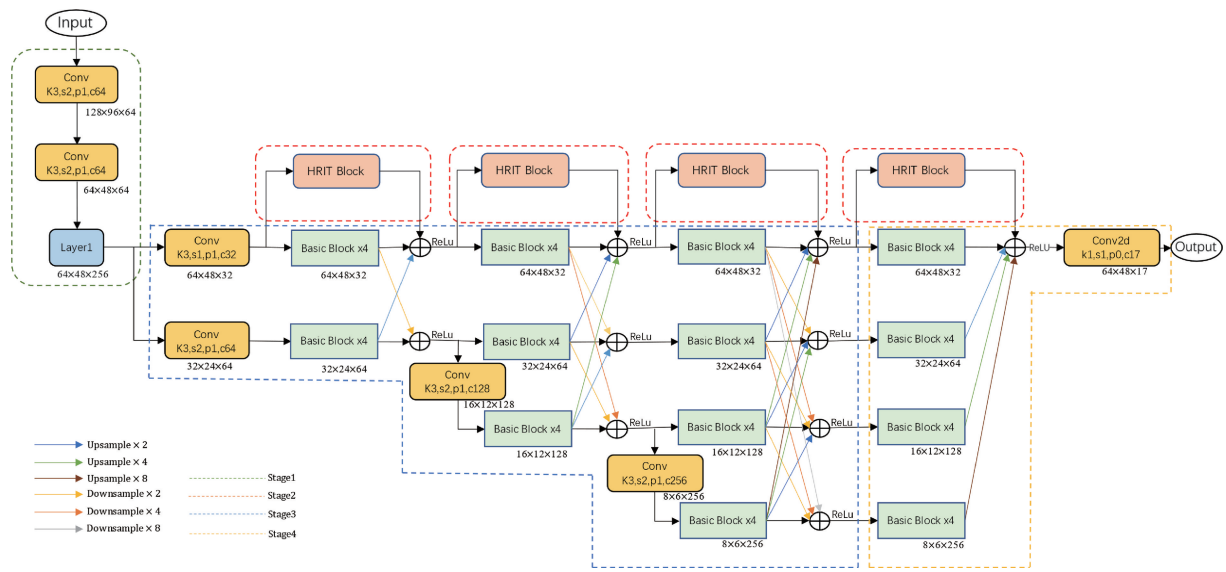


图 1 HRIT-HRNet 结构网络

融合使用。第 2 个阶段是信息的交互融合，在生成了不同分辨率的特征图之后，网络会引入一个过渡阶段，允许这些特征图相互融合。使用一组密集连接的跨分辨率残差块来处理特征图。每个残差块都包括跨分辨率的信息交互和特征融合操作，以及一个局部残差块，用于增强特征表示的非线性能力。第 3 阶段涉及在高分辨率特征提取网络上运行 HRIT 模块，使用 Transformer 自注意力机制提取高分辨特征图的全局特征，将其与主网络提取的局部特征进行前向输出融合。第 4 阶段是高分辨率提取阶段。特征图通过一组高分辨率卷积分支进行进一步处理，以提高分割或姿态估计的精度。

第 1 阶段为了构建包含不同尺度特征的特征图，在引出多分辨率分支之前，首先引用了 2 个 3×3 的卷积和 1 个 shuffle block 对图像进行 4 倍下采样。随后，在面经过交叉融合后的 3 个分支中，每个分支都由两个条件通道权重模块组成，通过深度可分离卷积进行特征提取。

第 2 阶段的信息交互融合中，根据不同分辨率的特征图的传递方向，采用上采样或下采样以及前向传输来实现。为了降低分辨率，进行 1 次步长为 2 的 3×3 卷积以实现 2 倍下采样；对于 4 倍分辨率降低，进行 2 次步长为 2 的 3×3 卷积以实现 4 倍下采样；同理，对于 8 倍分辨率降低，进行 3 次步长为 2 的 3×3 卷积以实现 8 倍下采样。若需要提高分辨率，则使用 1×1 卷积核，经过 BN 层，最后使用最邻近插值方法将其放大 n 倍，以获得上采样 n 倍后的结果。在完成分辨率交换后，将其他分支的特征图上采样或下采样以获得相同分辨率的特征图，并将它们相加，然后通过 ReLU 函数融合输出。此外，每次交换阶段中，分辨率最低的特征图通过一个步长为 2 的 3×3 卷积的下采样引出一个额外的特征图。对于输入为 $\{X_1, X_2, \cdots, X_s\}$ 输出为 $\{Y_1, Y_2, \cdots, Y_s\}$，交互公式为：

$$Y_k = \sum_{i=1}^{k} a(X_i, k) \tag{1}$$

若是进行跨分辨率之间交换则是具有额外的输出映射，公式为：

$$Y_{s+1} = a(Y_s, s+1) \tag{2}$$

式中：k 代表融合尺度个数；$a(,)$ 仅代表一个融合标识符。当其中 $k = 1$ 时，$Y_k = \sum_{i=1}^{k} a(X_i, k) = X_i$。

第 3 阶段在模块的输入中，将特征图 $X \in R^{N\times D}$ 分割成一系列不重叠的区间 $X \rightarrow \{X_1, X_2, \cdots, X_m\}$，其中每一块区间的大小均为 $P \times P$，并且在每一块区间上分别行使多头自注意力机制(muti-head self-attention，MHSA)。这种情况下网络可以更准确高效的提取每一块区间上的局部特征，具体框架如图 2 所示。第 m 块区间上的 MHSA 公式为：

$$Head(X_m) = Concat[head(X_m)_1, \cdots, head(X_m)_H]\boldsymbol{R}^{(K^2\times D)} \tag{3}$$

$$head(X_m)_h = Softmax\left[\frac{(X_m\boldsymbol{W}_q^h)(X_m\boldsymbol{W}_k^h)^{\mathrm{T}}}{\sqrt{D/H}}\right]X_m\boldsymbol{W}_v^h \in \boldsymbol{R}^{(K^2\times\frac{D}{H})} \tag{4}$$

$$\hat{X}_m = X_m + Head(X_m)\boldsymbol{W}_o \in \boldsymbol{R}^{(K^2\times\frac{D}{H})} \tag{5}$$

式中：$\boldsymbol{W}_o \in \boldsymbol{R}^{D\times D}$，$\boldsymbol{W}_q^h \in \boldsymbol{R}^{\frac{D}{H}\times D}$，$\boldsymbol{W}_k^h \in \boldsymbol{R}^{\frac{D}{H}\times D}$，$\boldsymbol{W}_v^h \in \boldsymbol{R}^{\frac{D}{H}\times D}$。$H$ 表示检测头的数量，D 代表检测通道的数量，N 表示输入图像的分辨率，qkv 则是执行注意力机制经过线性变换后得到的矩阵，注意力机制结构如图 3 所示，通过本身序列提供的 q 结合其他序列所提供的 k 得到一系列权重 α。再将 α 与 v 的乘法进行加权操作即可得到输出。其中 $\hat{X}_m$ 代表经过多头自注意力机制的输出结果。在将 $\hat{X}_1 \sim \hat{X}_m$ 相结合即可得到经过 MHSA 最后的输出。由于每个区间在特征提取时相互独立，因此在特征提取后，本文引入了一个前馈神经网络（feedforward neural network，FNN），在 FNN 两个多层感知机之间添加一个 3×3 的深度卷积层，以实现局部特征与全局特征的融合。这种方法有助于提高密集检测的性能。

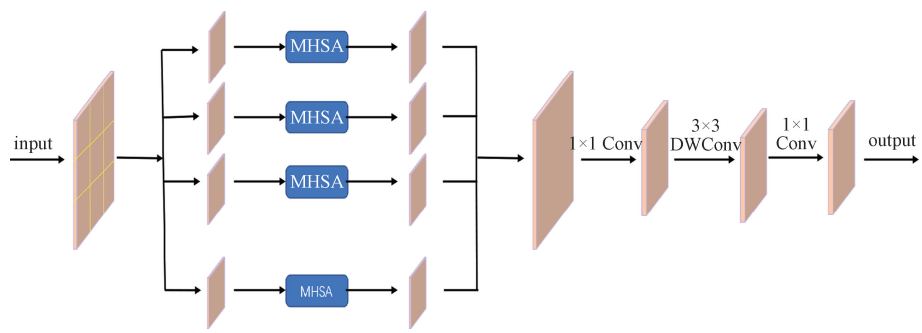


图 2　HRIT 模块内部结构

第 4 阶段的任务是在最终输出之前将所有分辨率的特征图融合成一张高分辨率的特征图。为实现这一目标，本文使用一个 1×1 大小的卷积核以及 BN 层，将不同分辨率的特征图放大 n 倍，然后进行融合。这个过程采用最邻近插值法来实现。

1.2　双通道物料大小判断执行机制

本文的算法旨在对电子元器件进行检测和定位，并计算它们的数量。经过测试，上述模型在十分密集的区域中

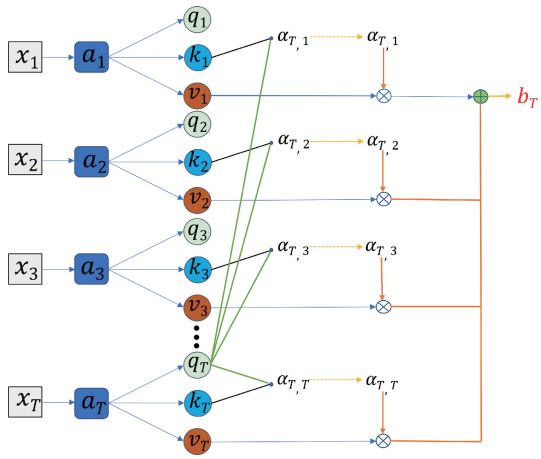


图 3　attention 机制结构

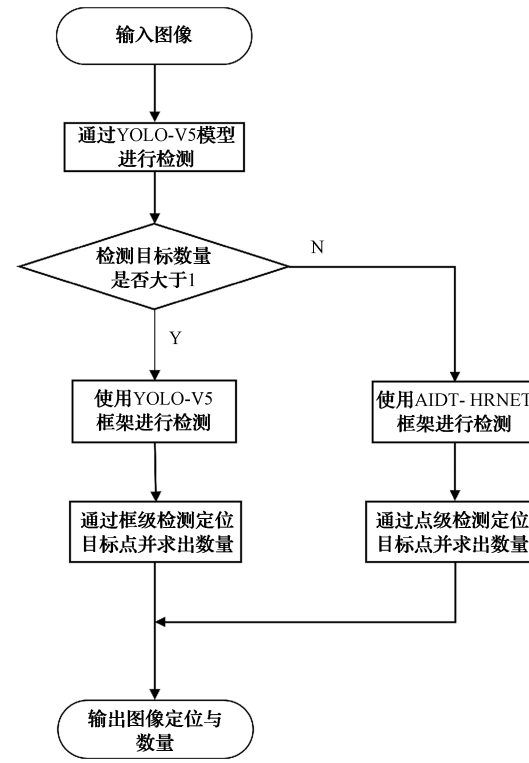


图 4　双通道物料大小判断执行机制流程

也能实现精确的检测。然而，对于较大的物料，可能会出现一次检测多个物料的情况，从而导致检测精度偏差。针对此种情况本文提出一种双通道物料大小判断执行机制，流程如图 4 所示。在进行检测之前，首先对输入图像进行判断。对于大块物料与密集物料，使用 YOLO v5 模型进行检测，细小密集元件会被检测为一个整体样本，因此输出为 1，而大块元件会被检测为一个单独样本，故输出大于 1。当输出大于 1 时，使用 YOLO v5 模型进行检测，将每个目标样本用正方形边框进行框选，并且输出边框中心点坐标与总数。否则，将继续通过 HRIT-HRNet 模型进行处理，输出每个样本的坐标和总数，然后将两者的结果汇总输出。

1.3　FIDTM 映射

FIDTM 映射是基于欧几里得映射（Euclidean distance transform map，EDTM）得到的针对密集细小物体更有效的映射图，公式如下：

$$P(x,y)=\min_{(x',y')\in B}\sqrt{(x-x')^2+(y-y')^2} \tag{6}$$

$$I=\frac{1}{P(x,y)^{(\alpha\times P(x,y)+\beta)}+C} \tag{7}$$

式中：B 代表了所有标注点的总和；(x,y) 代表一幅图像中的每一个像素点；(x',y') 代表 B 其中的某个标注点。$P(x,y)$ 表示图中每个像素点到标注点之间的最小距离。通过这种方法取得 $P(x,y)$ 范围，并且将其定义在 $[0,\infty]$。然而，由于范围过大，很难精确地进行映射回归。引出 FIDT 算法如式(7)所示。首先，通过取反加非零常数 C 将取值定义在 $[0,1/C]$。此时，将 C 置于 1 即可将范围定义在 $[0,1]$。通过这种方式，可以精确设置阈值并准确求出局部极大值的位置，实现较好的定位功能。

本文使用 HRIT-HRNet 网络对 FIDTM 进行回归，因为 FIDTM 映射上的每个点都对应一个局部最大值，所

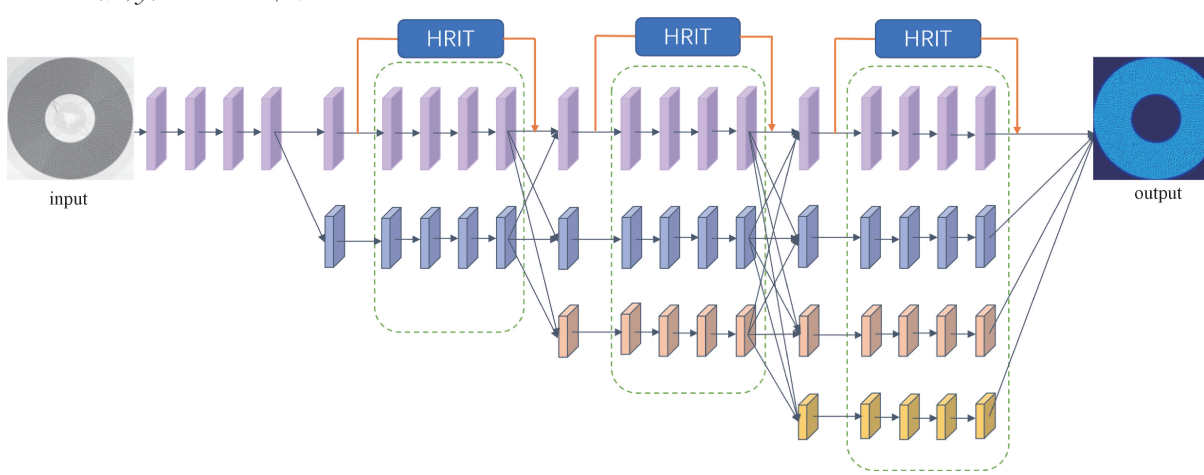


图 5　FIDTM 映射流程

以需要高分辨率的输出，本文网络恰好符合需求。输入图像经过回归输出 FIDTM 映射过程如图 5 所示。

1.4 损失函数

本文损失函数如下：

$$SSIM(E,G)=\frac{(2\mu_E\mu_G+\lambda_1)(2\sigma_{EG}+\lambda_2)}{(\mu_E^2+\mu_G^2+\lambda_1)(\sigma_E^2+\sigma_G^2+\lambda_2)} \tag{8}$$

$$L_S(E,G)=1-SSIM(E,G) \tag{9}$$

$$L_{I-S}=\frac{1}{N}\sum_{n=1}^{N}L_S(E_n,G_n) \tag{10}$$

$$L=L_{MSE}+L_{I-S} \tag{11}$$

式中：E、G 代表着预测成像与真实成像；μ 和 σ 代表平均值和方差。为了防止 μ 和 σ 为 0 出现分母为 0 的情况取常数 φ_1 和 φ_2 设为 0.000 1 和 0.000 9。此种情况下 $SSIM(E,G)$ 的大小将会被限制为 $[-1,1]$，值取到 1 时表示此时预测成像与真实成像相同，针对这种情况引出式(9)，此时 $L_S=0$ 代表效果最好。因为本文任务属于密集检测，所以需要更加重点地提取局部特征信息。式(10)较好地实现了这一需求，其中 N 代表了物料样本的总数，n 代表了将输入图像分成了 n 块 patch，每个 patch 距离设置为 30×30，E_n 与 G_n 分别代表了该块区域物料样本的预测数值与真实数值，通过这种方法可以使模型更加专注于局部区域，对于提取局部极大值和背景区域能力有较好的提升，最终与平均绝对误差(mean absoulte error，MAE)损失函数相加即得到本文损失函数。

1.5 局部极大值检测策略

通过得到 FIDTM 映射可以通过定位其中的局部最大值来获得其中物料样本的具体位置，流程如图 6 所示。

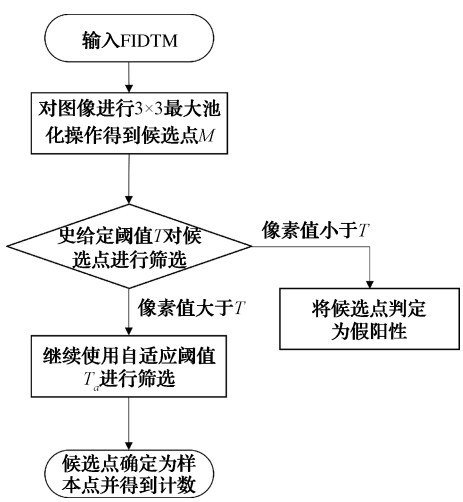


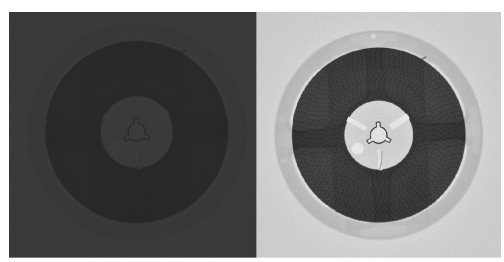
图 6 局部极大值检测流程

首先，使用 3×3 最大池化对整个映射进行筛选以获得新的样本点，这些点被预测为候选物料样本点。然而，在候选样本点中，由于背景的影响，仍然可能存在大量的假阳性，而实际的真阳性像素值通常要远大于假阳性。因此设置了一个阈值 T 用于筛选真阳性，如果经过筛选的一系列候选点 M 的局部最大值小于 T，则被视为假阳性并且筛选掉。如果大于 T，则进一步使用自适应阈值 T_a 对其进筛选，其中 T_a 值为候选点中 M 局部最大值的 100/255 倍。如果局部最大值大于这个阈值，则被视为真阳性并且计数加一，否则被视为假阳性并筛选掉。最后，将计数结果相加以得到统计数量。

2 实验结果与分析

2.1 数据集与训练环境

本文使用数据集皆为 X 射线无损检测[19-20]设备对 4 400 盘物料进行采集，将物料盘水平放入承载滑柜将其推入检测设备内，设备滑柜四周均设有光栅防护系统防止射线能量外泄，滑柜上方中心的 X 光射线源会由上至下进行放射成像，针对不同大小的物料盘分别进行测量，得到分辨率为 3 072×3 072 的 16 位 TIF 图像。所采集物料均来自华为、比亚迪、富士康等企业合法取得。其中训练集 3 080 张，验证集 880 张，测试集 440 张。由于采集图像目标点较小，故每张图片均对其中样本点中心进行点级标注。检测目标从 0.1～5 cm 不等。包括二极管、三极管、LDO 器件等微小半导体。图 7(a)为经过 X 射线检测设备采集到的样本图像，图 7(b)是对采集图像进行预处理后得到的 8 位图像。图 8 是对经过预处理后的图像样本点进行标注的面板。数据集存在如下难点：1)数据集中样本复杂多样，大小不等；2)包含多种样本较为密集的图像，且疏密程度分布不均匀。



(a) (b)

图 7 设备取图原图与经过预处理图像

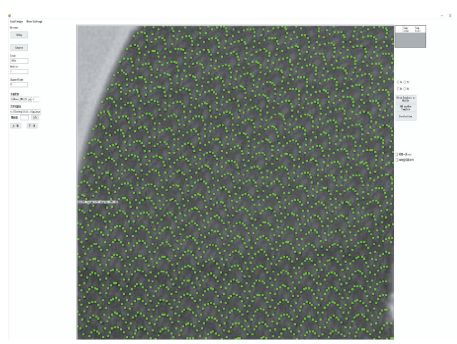


图 8 数据集标注界面

2.2 实验评估指标

为了评估本文算法的有效性，选取了精确率(preci-

sion，P)、召回率(recall，R)和综合评价指标(F1-score，F_1)来对检测定位进行评估，通过均方误差(mean square error，MSE)与 MAE 对计数进行评估，检测速度用 s 为单位的推理时间 T 来表示。其中各项指标表达式如下：

$$P = \frac{TP}{TP + FP} \tag{12}$$

$$R = \frac{TP}{TP + FN} \tag{13}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{14}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2 \tag{15}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | y_i - y_i' | \tag{16}$$

式中：TP 代表预测为正类的正样本；FP 代表预测为正类的负样本；FN 代表预测为负类的正样本；n 表示数据集数量；y_i 表示第 i 张图像数量的真实值；y'_i 表示第 i 张图像数量的预测值。

实验训练使用系统为 Ubuntu 20.04，所用来训练的平台 GPU 为 6 块 NVIDIA GeForce RTX 3060 搭载的服务器，内存大小为 72 GB，torch 版本为 1.8.0。

2.3 实验参数设置

为了提高数据集的训练效果，采用了随机裁剪和水平翻转等数据增强技术。输入图像大小为 3 072×3 072，将其裁剪为 512×512。FIDT 中 α 与 β 分别取 0.02 与 0.75，学习率设置为 0.000 1，训练批次设置为 200。

2.4 消融实验与分析

为了验证模型改进的有效性，本文在以 HRNet 模型为基础进行实验。针对第 1 个创新点，实验 1 为原有网络，消融实验 2 在 HRNet 将高分辨率交互部分替换为 HRIT block，消融实验 3 在 HRNet 高分辨率交互模块并联 HRIT block。针对第 2 个创新点，消融实验 4 在 HRNet 网络外附加运行 YOLO V5 网络列出以下组合模型，消融实验 5 即为本文模型。得到结果如表 1 所示。

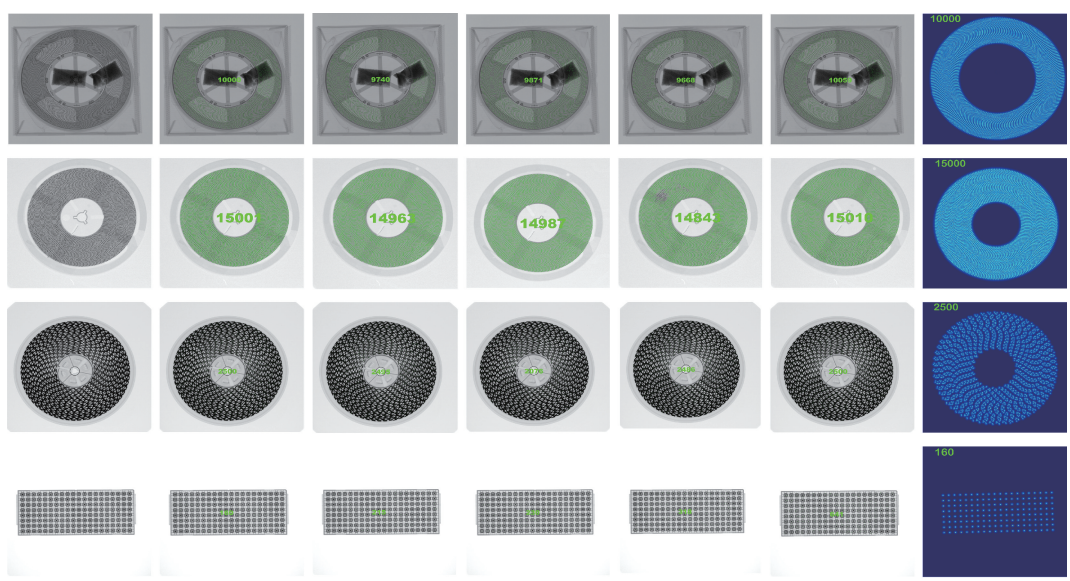
表 1 不同模块消融实验对比结果

实验	MSE	MAE	F_1/%	P/%	R/%	T
1	576.4	80.2	84.1	82.9	85.3	**104.3**
2	485.1	93.6	82.4	90.9	92.0	108.2
3	249.4	40.2	94.3	93.5	95.3	110.9
4	552.4	73.7	87.4	87.6	93.5	116.4
5	**186.3**	**32.6**	**96.7**	**95.3**	**98.2**	124.1

通过消融实验对比可以看出，实验 2 在串联了自注意力机制模块后，虽然其他指标有少许提升但综合评价指标下降 1.7%，实验 3 在并联了自注意力机制模块后各项指标均有明显提升，综合评价指标较实验 1 上升 10.2%；实验 4 在简要结合了 YOLO v5 网络共同检测后，相较实验 1 各项指标也都有提升。于是本文算法针对实验 3 与实验 4 优化进行改进，在主干网络并联子注意力机制模块的同时采取双通道进行判断，使得综合评价指标相较于原网络提升 12.6%，可以有效识别图像特征进而对其进行精准的定位与计数，满足了工业上对精密电子元件检测计数的要求。

2.5 对比试验分析

为了能准确地分析出本文模型在电子物料检测方面的可行性，本文选择 LCS-CNN[21]、Topocount[22]、Crowd-SDNet[23]、AutoScale[24] 4 种模型与本文模型进行对比分析。通过参数量、运算时间、平均绝对误差、均方误差来对比分析结果如表 2 所示。



(a) **样本原图** (b) **本文算法** (c) LCS-CNN (d) Topocoount (e) Crowd-SDNet (f) AutoScale (g) Ground truth

图 9 不同算法检测结果

表 2 不同算法检测结果

方法	MSE	MAE	F_1/%	P/%	R/%	T
LCS-CNN	779.5	196.3	66.8	69.7	64.2	**107.2**
Topocount	381.3	73.2	87.1	86.3	88.0	113.3
Crowd-SDNet	266.5	70.8	89.7	90.1	89.5	120.5
AutoScale	247.9	45.4	92.6	91.6	93.7	143.2
本文	**186.3**	**32.6**	**96.7**	**95.3**	**98.2**	124.1

可以观察到，相对于其他模型，该模型在牺牲少量时间的情况下能够显著提高检测精确度，从而获得最精准的计数结果。证实本文所提出模型在工业检测上对电子物料上有较好的识别能力。

通过实验结果对比可以得知，本文算法在所有对比实验算法中的综合评价指标达到了 96.7%，远超于其他算法的检测精度，进而可以很好地实现对样本点的精确定位。并且本算法的 MAE、MSE 均保持最低，保证了计数的精准性。本文算法与其他算法针对不同特征物料成像对比如图 9 所示，可以看出本文算法不论检测样本在物料过大过小情况下都能做到精准检测，并且在遮挡、粘连的情况下也可以进行精确的定位与计数。

3 结 论

本文在 HRNet 网络的基础上提出了 HRIT-HRNet 网络，用于解决原网络在样本点数量过多且分布不均匀的情况下无法准确定位和计数的问题。原网络只适用于样本点数量较少且分布均匀的场景，而在挑战更大的情况下效果较差。本文提出一种双通道自注意力多尺度融合机制模型，在主干网络的高分辨率通道上额外添加了自注意力机制，同时兼顾提取了全局信息与局部信息，改善了原网络局部特征信息提取不足的问题。此外，此外，在图像输入阶段，对图像进行初步检测和分类，有助于更精确地进行计数。尽管本文模型在检测和计数方面取得了良好的效果，但仍有改进的空间。添加了自注意力机制虽然增强了局部特征提取的能力，但是相应也会增加检测所需时间，在后期的研究工作中，可以针对此方向对其进行改良以此提升检测效率。

参 考 文 献

[1] DANIEL O R，ROBERTO L S. Towards perspective-free object counting with deep learning[C]. European Conference on Computer Vision，2016：615-629.

[2] 李杰. 基于单阶段网络的小目标检测算法研究[D]. 成都：电子科技大学，2023.

[3] 于波. 基于 YOLOv5 的双阶段小样本目标检测研究[D]. 南昌：华东交通大学，2023.

[4] ROSS G，JEFF D，TREVOR D，et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. Proceedings of the IEEE Conference on Computer Vision andPattern Recognition，2014：580-587.

[5] ROSS G. Fast R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision，2015：1440-1448.

[6] KIM B，CHO S. Image-based concrete crack assessment using mask and region-based convolutional neural network[J]. Structural Control and Health Monitoring，2019，26(8)：23-81.

[7] LIN T Y，PIOTR D，ROSS G，et al. Feature pyramid networksfor object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition，2017：2117-2125.

[8] SHANG C，AI H Z，BAI B，End-to-end crowd counting via joint learning local and global count[C]. International Conference on Image Processing，2016：1215-1219.

[9] ZHANG S，WU G，COSTEIRA J P，et al. FCN-rLSTM：Deep spatio-temporal neural networks for vehicle counting in city cameras[C]. International Conference on Computer Vision，2017：3687-3696.

[10] 孙备，党昭洋，吴鹏，等. 多尺度互交叉注意力改进的单无人机对地伪装目标检测定位方法[J]. 仪器仪表学报，2023，44(6)：54-65.

[11] HOSSAIN M，HOSSEINZADEH M，CHANDA O，et al. Crowd counting using scale-aware attention networks[C]. Winter Conference on Applications of Computer Vision，2019：1280-1288.

[12] 蒋妮，周海洋，余飞鸿. 基于计算机视觉的目标计数方法综述[J]. 激光与光电子学进展，2021，58(14)：1400002.

[13] SUN K，XIAO B，LIU D，et al. Deep high-resolution representation learning for human pose estimation[C]. Conference on Computer Vision and Pattern Recognition，2019：5686-5696.

[14] 徐健，陆珍，刘秀平，等. 注意力机制优化 RetinaNet 的密集工件检测方法研究[J]. 电子测量与仪器学报，2022，36(1)：227-235.

[15] LIANG D K，XU W，ZHOU Y，et al. Focal inverse distance transform maps for crowd localization[J]. IEEE Transactions on Multimedia，2023，25：6040-6052.

[16] FU R，CHEN X L，HUANG L，et al. HRFormer：High-resolution transformer for dense prediction[C]. NeurIPS 2021，2021.

[17] 王立刚，张志佳，贺继昌，等. 基于深度学习的交通标志文字信息检测与识别方法[J]. 电子测量技术，2022，45(18)：119-125.

[18] 高强，唐福兴，李栋，等. 基于改进 YOLO v5 的密集场

景行人检测方法研究[J]. 国外电子测量技术,2023,42(4):125-130.

[19] 李俊霖,杨晨菲. X射线无损检测系统信息化技术的研究与应用[J]. 无损探伤,2023,47(2):38-41.

[20] 丁卫良,常华峰,潘龙龙,等. X射线无损检测的应用及发展趋势[J]. 科技创新与应用,2020(36):161-162.

[21] SAM D B, PERI S V, SUNDARARAMAN M N, et al. Locate, size and count: Accurately resolving people in dense crowds via detection[C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020:2739-2751.

[22] SHAHIRA A, MINH H, DIMITRIS S, et al. Localization in the crowd with topological constraints[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021,35(2): 872-881.

[23] WANG Y, HOU J H, HOU X Y, et al. A self-training approach for point-supervised object detection and counting in crowds[J]. IEEE Transactions on Image Processing, 2021,30:2876-2887.

[24] XU C F, LIANG D K, XU Y X, et al. AutoScale: Learning to scale for crowd counting and Localization[J]. International Journal of Computer Vision, 2019, DOI:10.48550/arXiv.1912.09632.

作者简介

朱希(通信作者),硕士研究生,主要研究方向为目标检测、语义分割、无损检测。

E-mail:1244659059@qq.com

李燕,博士,教授,主要研究方向为深度学习、智能计算。

E-mail:002200@nuist.edu.cn

施林枫,硕士研究生,主要研究方向为人脸识别、缺陷检测、遥感图像处理。

E-mail:1392906437@qq.com