对电子垃圾进行图像分割和分类，以训练机器人进行废物分类。

Prakriti Tripathi
Dept. of CSE
IIT Dharwad
Dharwad, India

Theertha Biju
Dept. of ME
IIT Dharwad
Dharwad, India

Maniram Thota
Dept. of ME
IIT Dharwad
Dharwad, India

Prof. Rakesh Lingam
Dept. of ME
IIT Dharwad
Dharwad, India

***Abstract*—行业合作伙伴提供了一个问题陈述，涉及使用机器学习模型对电子废物进行分类，该模型将由拾放机器人用于废物分拣。我们首先取常见的电子废物物品，如鼠标和充电器，将其拆焊，并拍照以创建自定义数据集。然后训练并运行最先进的 YOLOv11 模型 [?] ，以在实时中实现 70 的 mAP。Mask-RCNN 模型也经过训练并实现了 41 的 mAP。该模型将进一步与拾放机器人集成，以执行电子废物的分拣。代码和数据集的链接可以在此 github 存储库中找到： https://github.com/prakriti16/Image-segmentation-and-classification-of-e-waste**

电子废弃物（e-waste）是全球增长最快的固体废物流之一。在 2022 年全球产生的电子废弃物中，已知正式回收的不到四分之一；然而，如果电子废弃物流得到适当回收，则其中含有的宝贵且稀缺的资源可以被再利用。

手工清理和回收电子废物有害，因为其中含有铅等多种有毒物质。因此，我们研究使用机器学习模型和机器人来实现这一过程的自动化。

在工厂中的回收厂内，电子废物被粉碎后，诸如电阻器、电容器和 LED 等几个内部组件原样出现。

这些组件可以通过我们训练的模型识别出来，然后沿传送带的拾取和放置机器人可以根据大小或类型将这些物品分拣到不同的箱子中。

I. 文献综述

在 MS COCO 数据集上，一个流行的实时实例分割模型是基于 YOLOv5 模型的 RTMDet [?] ，该模型在 10.18 M 参数下实现了 44.0 的框 AP 和 38.7 的掩码 AP。

最近，Ultralytics [?] 发布了 YOLOv11 模型，该模型在具有 10.1 M 参数的情况下实现了 46.6 的 box AP 和 37.8 的 mask AP。

特别是在垃圾检测方面，Mindy Yang 和 Gary Thung [?] 尝试了 SVM 和 CNN 模型，在他们的定制垃圾数据集上实现了 75 % 的准确率，该数据集包含 2400 张图像，分为 5 类：纸张、纸板、金属、玻璃和其他。然而，这只是一个分类任务，而不是实例分割任务。

Pedro F. Proença 和 Pedro Simões [?] 使用了 Mask-RCNN，它在他们开发的语境垃圾标注数据集上具有 ResNet 和 FPN 骨干，该数据集包含自然和户外的垃圾图片，并有多个类别，如透明塑料瓶、玻璃瓶、电池等，这些被归类为瓶子等超级类别。他们仅实现了 26.1 的低 mAP，因为由于图像缩小后错过了多实例的小物体，如香烟，在分类中被遗漏。

尼富尔 · 伊斯兰等人 [?] 引入了 EWasteNet，这是一种数据高效的图像转换器，只需一百张图像即可进行训练。然而，由于其包含两个步骤的过程，因此需要额外的训练时间，使其不适合用于实时应用。

我们专门在电子废物的背景下创建了自己的数据集。大多数电子废物数据集包含处于崭新且无损状态的电子设备的图片。然而，在回收工厂中，经过粉碎处理后，这些物品将被打碎成小组件。因此，我们取了一些电子设备如鼠标和充电器，去除了外壳并拆焊了内部的各个组件。这模拟了电子物品被粉碎后的情况。我们使用了 YOLOv11 模型，并在对数据集进行预处理后对其进行了微调，达到了平均 70.6 的 box mAP50 和 70.8 的 mask mAP50，每张图像的处理速度为 11 毫秒。我们还尝试使用 Mask-RCNN 模型 [?] ，达到了 41.392 的分割 mAP50 和 42.293 的 bbox mAP50。

II. 方法论

A. 数据集

鼠标和充电器的内部组件被拆焊。对所有可能方向的单个组件拍摄了 100 张图像。然后录制了一段 22 秒的视频，其中所有组件被放置在彼此附近，呈一条直线排列，以模拟回收设施中皮带传动的视角。该视频以每秒 60 帧的速度采样，将单独帧的图像加入数据集。为了增加数据集的多样性，从互联网上增加了这些组件的另外 100 张图像。这样总共得到 643 张图像。

接下来，使用在线工具对图像进行了标注，每种类型的组件都被分配了一个类别。

预处理包括应用自动定向，以确保没有错位。图像被分成 2x2 的切片，以确保捕捉到小组件和细节。

数据增强包括水平和垂直翻转，以及旋转 90 度，并向像素添加 0.1 % 的噪声扰动以模拟相机质量。

最终数据集包含 6180 张图像，其中有 88 % 用于训练，8 % 用于验证，4 % 用于测试。

这些图像通过了 YOLO11s-seg 模型，该模型包含 355 层，10,089,254 个参数，10,089,238 个梯度，35.6 GFLOPs。

该架构涉及通过缩小空间维度（通过步长为 2 的卷积）对图像进行初始顺序下采样，而后面的行通过上采样和拼接来恢复空间分辨率。中间也有一些层，如 C3k2、SPPF 和 C2PSA，用于高效特征提取和注意机制。

C3k2：结合卷积层和瓶颈层进行特征提取。

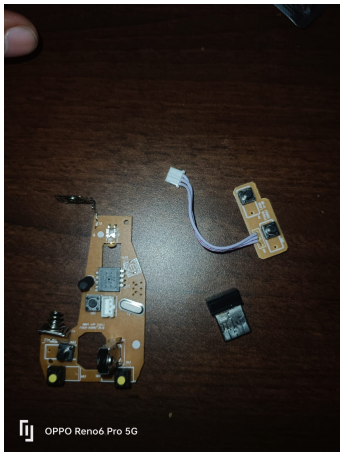

Fig. 1. 鼠标的内部组件。

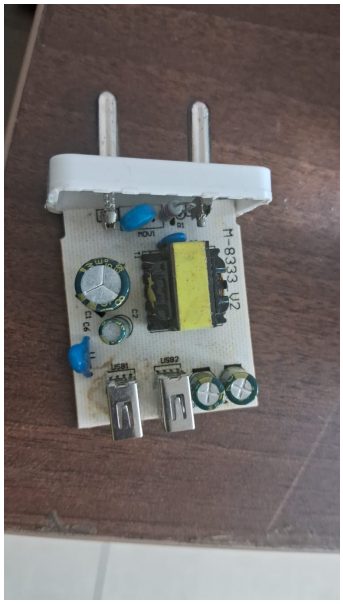

Fig. 2. 充电器的内部组件。

SPPF: 空间金字塔池化–快速。用于在多个尺度上进行特征聚合。

C2PSA: 基于注意力的特征优化。

上采样：将特征图的空间维度扩大至两倍。

Concat: 结合来自不同层的特征图。

B. Mask-RCNN 模型架构

Mask-RCNN 由 ResNet 主干网络组成，用于提取特征，其中深层捕捉语义信息，浅层捕捉细节。FPN 通过创建特征金字塔增强了主干网络，使得能够检测不同尺度的物体。然后这些通过区域提议网络，最后通过掩码预测头生成最终输出。

ResNet 层块：

例如：res5 的结构：第一个 BottleneckBlock：

快捷路径：一个 1x1 卷积将输入从 1024 通道投影到 2048 通道，同时将空间维度减半（步幅为 (2, 2)）。一个层用于规范化激活函数。

Fig. 3. 来自数据集的样本图像。它包含鼠标和充电器的几个小部件，并通过手机拍摄。

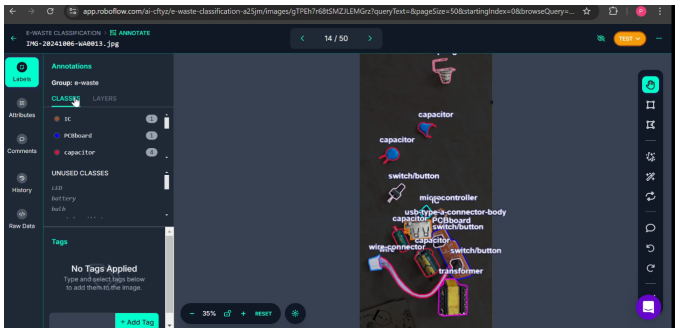

Fig. 4. 数据集中带注释的示例图像。

主路径：Conv1：一个 1x1 卷积，用于将通道数从 1024 增加到 2048。

Conv2：用于特征提取的 3x3 卷积。

Conv3: 一个 1x1 的卷积，用于将下采样后的特征图投影回到 2048 个通道。

每个卷积之后都会进行归一化。

RPN 生成潜在的感兴趣区域（ROI）以供进一步处理，其结构类似于 ResNet BottleneckBlock 的主路径。

ROI 头部通过全连接层细化区域提议并进行输出预测。

III. 结果与讨论

Mask-RCNN 模型的训练时间是 YOLOv11 的 3 倍，并且无法识别像鼠标滚轮编码器这样实例较少的类别中的对象。

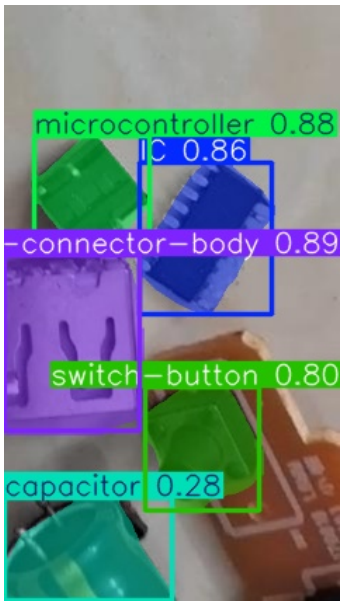

Fig. 5. YOLOv11 模型生成的分割掩码的示例输出。

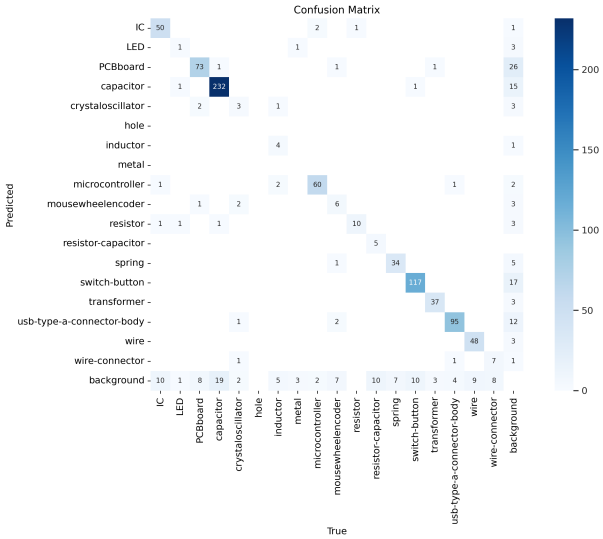

Fig. 6. YOLOv11 的混淆矩阵。

IV. 结论与未来工作展望

首先，通过拍摄鼠标和充电器未焊接组件的照片和视频创建了数据集。此外，还包括了 100 张来自互联网的图片。通过应用增强和预处理技术，数据集的大小从 643 增加到 6180。

其次，应用并比较了两个模型：YOLOv11 和 Mask-RCNN，针对各种指标进行了评估。

Model	R	mAP50	mAP50:95
YOLOv11	65.5	71.6	58.5
Mask-RCNN	38.4	41.1	26.9

TABLE I
两种模型在不同指标和边界框交并比上的比较。

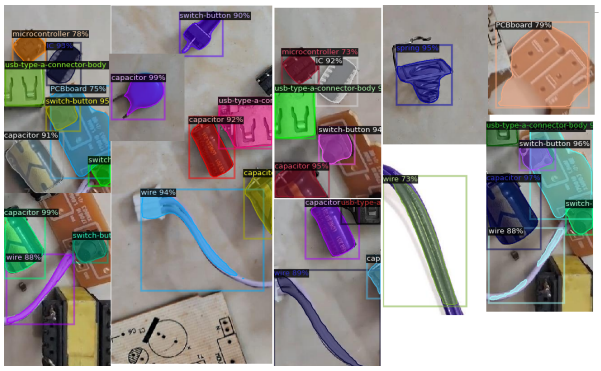

Fig. 7. 来自 mask-RCNN 模型的示例输出。

Model	R	mAP50	mAP50:95
YOLOv11	65	70.7	51.6
Mask-RCNN	36.1	39.7	23.7

TABLE II
两个模型在各种指标和遮罩分割交并比上的比较。

YOLOv11 大约需要 1 小时进行训练，而 Mask-RCNN 则需要 3 小时。因此，我们可以得出结论，YOLOv11 在所有指标上均优于 Mask-RCNN。

我们可以在将来扩展这个数据集，以涵盖其他类别的电子废弃物设备。同时，我们计划将这个计算机视觉模型与机器人软件集成，用于工业中的实际应用。需要在回收设施中创建一个详尽的电子废弃物数据集，当这些废弃物被粉碎后，以便我们可以创建模型有效地进行分拣。