arXiv:2305.04151v2 [cs.CV] 8 Sep 2023

上下文感知图表元素检测

Pengyu Yan*、Saleem Ahmed* 和 David Doermann

计算机科学与工程系，布法罗大学，布法罗，纽约州{pyan4、sahmed9、doermann}@buffalo.edu

摘要。 作为图表数据提取的前提，准确检测图表基本元素至关重要且不可或缺。与通用图像领域中的目标检测不同，图表元素检测高度依赖上下文信息，因为图表是高度结构化的数据可视化格式。为此，我们提出了一种新方法 **CACHED**，即 **上**下文**感**知**图**表 **元**素 **检**测，通过将由视觉上下文增强和位置上下文编码组成的局部-全局上下文融合模块与 Cascade R-CNN 框架相结合来实现。为了提升方法的泛化能力，使其适用于更广泛的应用场景，我们改进了现有的图表元素分类，并将图表基本元素标准化为 18 个类别，不包括绘图元素。在更新图表元素类别后，我们的 CACHED 方法在实验中取得了最先进的性能，凸显了上下文在图表元素检测中的重要性。将该方法扩展到柱状图检测任务后，我们在 PMC 测试数据集上取得了最佳结果。我们的代码和模型可在 https://github.com/pengyu965/ChartDete 获取。

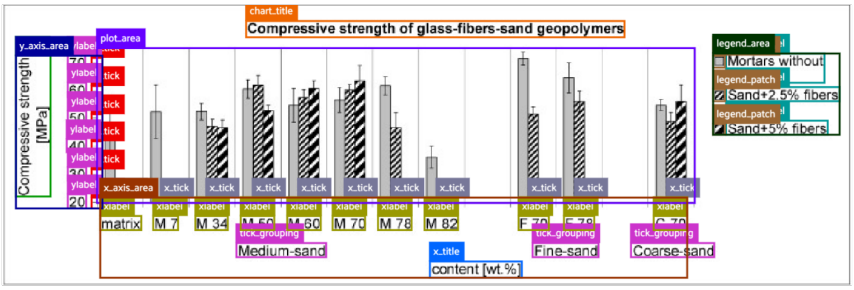
关键词： 图表检测 · 图表数据提取 · 图表理解 · 文档分析

1 引言

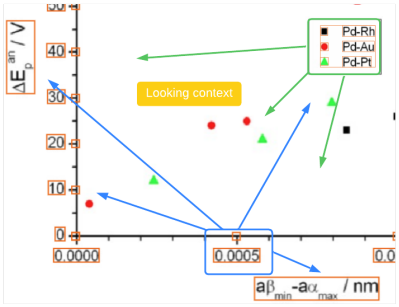
图表是高度抽象的数据可视化格式，便于读者获取不同实体之间的趋势或比较信息，但很难从中提取确切的数据值。因此，自动化图表数据提取可以减少人们从科学、金融分析、营销图表等内容中总结数据所需的工作量。图表数据提取通常涉及但不限于元素检测、文本光学字符识别、数据解读和语义数据转换。基本元素检测是图表数据提取中最基础的部分，并会影响所有下游任务。因此，准确检测和识别绘图区外的图表基本元素（见图 1）是至关重要的第一步。由于图表设计高度多样，基本元素检测具有很大挑战性。

已提出多种方法 [24,25] 用于数据提取，但仅关注数据绘图检测，例如柱状图和折线关键点，却忽略了大多数基础

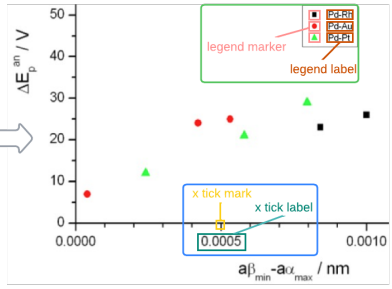
* 共同贡献



(a) Chart Element Detection Objective



(b) Detection with Local-Global Context



(c) Detected Samples

图1：**上下文感知图表元素检测**。(a) 是图表元素检测目标的示例，(b) 和 (c) 展示了图表图像中的检测依赖于上下文。

元素检测。同时，一些研究 [26,1,12] 常使用常见的标准两阶段检测器来检测图表元素。这些两阶段检测器 [10,9,30,2,13] 利用图像上下文的能力有限，而上下文对于在图表图像领域实现准确检测至关重要。与一般图像中的常见对象不同，图表图像中的许多元素在视觉上相似，但作用各异。只有参照上下文才能加以区分，例如，图例标签和刻度标签都是文本块，但它们在图表图像中的功能位置以及与其他元素的关系不同，因此所起的作用也不同（如图1(b)(c)所示）。因此，需要一种利用局部-全局上下文特征的检测器来解决图表元素检测任务。此外，对图表元素进行全面且合理的分类，有助于提升模型在各类图表上的泛化能力。

本文提出一种上下文感知图表元素检测方法，在 Cascade R-CNN 的级联 RoI 头之间集成局部—全局上下文融合模块，以凸显上下文在图表中的重要性。如图2 所示，局部—全局上下文融合模块包含两部分：视觉上下文增强（VCE）

和位置上下文编码（PCE）。VCE 将全局特征图融入局部特征图，使模型能够获取更丰富的视觉上下文信息。PCE 则使模型能够根据边界框坐标学习特定的对象分布模式。此外，我们研究了多个数据集，以分析并完善图表元素分类。我们总结了图表元素的 18 个类别，不包括任何绘图元素，并相应更新数据集，用于模型训练和测试。定量评估和样本定性分析表明，我们的方法能够准确检测图表元素。**总体而言，我们的贡献可总结如下：**

– 我们提出了一种采用局部-全局上下文融合模块的检测器，以实现精准的图表元素检测。该模块从视觉特征和位置特征两个方面强调上下文信息。我们的方法在图表元素检测任务上取得了最先进的结果。– 改进图表元素分类，并生成额外的结构区域对象，以帮助检测器更好地理解图表。我们共总结出 18 个类别，更新后的 PMC 数据集可通过以下地址获取：
https://github.com/pengyu965/ChartDete– 我们的方法以及若干常见的两阶段检测器（包括现有相关研究中使用的检测器）均在更新后的数据集上进行了训练和评估。这有助于全面了解这些方法在此类任务上的表现。

2 相关工作

2.1 目标检测

自 2014 年以来，许多设计精良的目标检测器相继问世，大多数可分为两类：单阶段检测器 [22,27,28,29] 和两阶段检测器 [10,9,30,2,13]。它们的第一部分相似，均采用基于卷积神经网络（CNN）的骨干网络 [14,35,8,15,36,16,31,32] 来提取视觉特征。两者的区别在于，两阶段检测器包含一个区域提议模块，用于生成与类别无关的区域提议，然后在第二阶段对这些对象进行分类并细化其定位；而单阶段检测器则直接根据输入图像及其特征图预测对象的位置和类别。直观而言，Faster R-CNN [30] 和 Cascade R-CNN [2] 等两阶段检测器在目标定位和识别方面准确率更高，但推理速度较慢。图表元素检测任务对目标定位和分类准确率的要求高于对推理速度的要求，因此我们选择使用两阶段检测器来完成该任务。

2.2 Transformer

随着基于Transformer的方法 [34,7] 在自然语言处理领域取得成功，许多研究 [33,19,17] 致力于将Transformer应用于视觉相关任务。Transformer能够从视觉和文本输入中提取强大的特征表示。2020年，Carion等人 [3] 提出了首个基于Transformer的端到端目标检测器。然而，DETR在小规模数据集上的训练稳定性不足，且无法

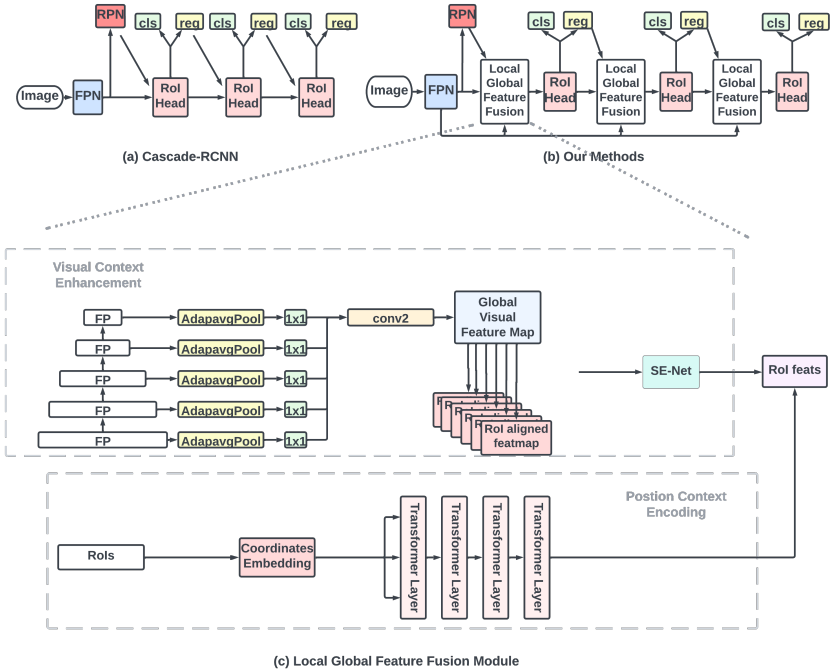


图2：**我们方法的框架**。(a) 为标准的 Cascade R-CNN 模型，(b) 为我们方法的整体框架，(c) 展示了我们方法中局部-全局特征融合模块的细节。

准确定位和识别小目标或重叠目标。在一项较新的研究中，[23] 提出了一种分层Transformer，利用移位窗口从输入图像中提取特征，所得框架可作为两阶段检测器的新骨干网络。研究表明，这一新骨干网络能够有效提升两阶段检测器的性能。由于Transformer架构中的注意力机制，Swin Transformer能够关注空间信息并获取具有上下文感知能力的图像特征。我们的模型和实验均专门使用Swin-Transformer 主干网络。

2.3 图表数据提取

2017年，Jung 等人 [26] 提出了一种半交互式系统，用于从图表中提取底层数据。然而，该系统首先需要人工交互来设置起点和终点。随后，系统采用基于规则的方法解读数据值。这种方法的数据提取高度依赖人工交互，在复杂情况下会失效。2019—2022年，Davila 等人 [4,5,6] 组织了图表采集竞赛，并提供了两个有价值的数据集：来自 PubMed Central 文献的人工标注的真实世界图表图像（PMC 数据集）和 Adobe 合成图表数据集。作为 2020年图表采集竞赛的参赛团队，Ma 等人 [25]使用

Cascade R-CNN 和基于热力图的关键点检测器分别用于检测条形框和折线关键点。该方法聚焦于绘图区内的数据绘图检测；在数据解读方面，则使用元素的真实标注与预测的数据绘图进行语义转换。之后，另一种 EXCEL400K 数据集在 [24] 中提出，该方法利用 Microsoft Excel API。与 CornerNet [18], 类似，该方法采用带有关键点热力图生成功能的沙漏网络骨干网络，预测并分组对象上的角点集合。作者利用关键点检测方法，将检测方法推广至不同类型的图表。然而，这项工作提出的数据集仅包含绘图区内的数据绘图标注，许多最基本且必不可少的元素都未加标注。

3 我们的方法

本节将从以下几个方面介绍我们的 CACHED 方法：局部-全局上下文融合模块，该模块由视觉上下文增强和位置上下文编码器组成；针对类别级对象不平衡的损失函数；以及适用于更广泛场景并具有更强泛化能力的标准化分类方法。

3.1 局部-全局上下文融合模块

如图2 所示，局部-全局上下文融合模块经过设计并集成在各个 RoI head 之间，用于上下文提取与融合。在将特征送入 RoI head 进行回归和分类之前，这些模块会为每个感兴趣区域提供局部-全局视觉特征和相对位置特征。这三个模块架构相同，但采用不同的权重进行训练，每个模块均由以下两个部分组成。

视觉上下文增强（VCE）尽管随着卷积神经网络（CNN）层的堆叠，每个锚点的视野会增大，但其仍然局限于局部范围，上下文感知能力较弱。然而，图表图像中的准确元素检测与标注需要更大的上下文范围（见图1）。例如，一个文本块可以根据几乎来自整幅图像的上下文被分类为图例标签：该图例标签位于一些（图例）标记旁边，而这些标记与绘图区中的绘图元素具有相同的颜色。为了解决这一问题，类似于 SCNet [11], ，我们通过将骨干网络特征金字塔中的特征图作为全局视觉特征，并将其引入每个感兴趣区域，提出了视觉上下文增强（VCE）。这些全局视觉特征与每个 RoI 对齐的局部特征图相结合，以增强局部—全局视觉上下文。如图 2(c) 所示，我们首先将特征金字塔所有阶段的特征图进行平均池化，使其具有相同的尺寸 $[N, 256, 7, 7]$，然后将它们拼接起来，此时其形状变为$[N, 256 * 5, 7, 7]$。卷积神经网络进一步抽象这些全局视觉特征，并减少通道数。随后，将全局特征图附加到每个 RoI 对齐的特征图上。每个尺寸为$[N, 512, 7, 7]$ 的感兴趣区域的特征图现在由前一半通道中的局部视觉特征和其余通道中的全局视觉特征组成。SE-Net 和一个 1x1 卷积层将

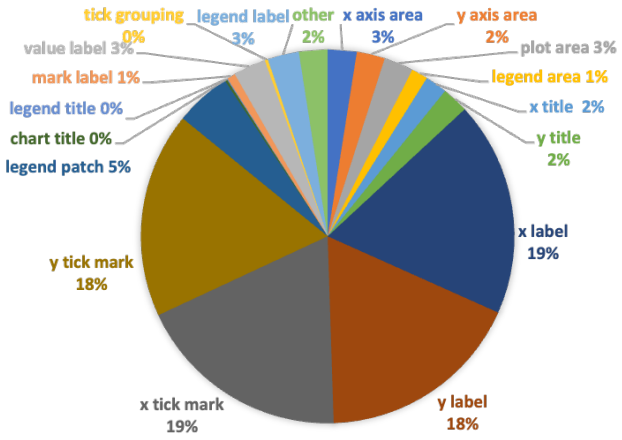


图3：**PMC 数据集中按类别划分的图表对象数量分布**。图表图像中的对象在类别上天然分布不均衡。x/y轴刻度标签和x/y轴刻度线的数量不可避免地多于其他对象。有些元素极小，例如刻度分组和图例标题。

分别用于按通道融合局部—全局视觉特征，并减少通道数。最终尺寸为 $[N, 256, 7, 7]$ 的特征图与原始 RoI 特征图的维度一致，但包含丰富的局部—全局视觉上下文，因此被用作每个感兴趣区域的视觉特征表示。

位置上下文编码（PCE）在通用图像领域，对人员、车辆和动物等常见对象的检测主要依赖视觉特征，而对象在图像中的位置对准确定位和分类的影响较小。然而，由于图表图像是高度结构化的数据可视化格式，对象的呈现始终遵循某些特定模式，例如，x轴刻度标签始终位于x轴刻度线下方或旁边，图表标题通常出现在图像边缘，而很少出现在中心区域，因此图表元素之间的相对位置上下文非常重要。我们提出了一种基于Transformer的位置上下文编码器，用于获取位置上下文特征。如图2 所示，基于Transformer的架构在对象的边界框坐标之间建立注意力。首先，我们将4维坐标归一化为 $[0, 1]$ ，归一化依据是输入图像的高度和宽度。随后，线性层将归一化坐标嵌入为512维向量。在嵌入边界框向量序列中填充零掩码填充，其中最终嵌入边界框序列的长度固定为1024。1024的块大小足以覆盖训练和测试过程中RPN结果中的最大采样边界框数量，该数量最多设为1000。我们使用每个注意力头的输出表示从对应边界框到所有其他边界框的相对位置上下文信息。然后，将每个边界框的编码向量与其RoI视觉特征向量进行拼接。

3.2 类别间对象数量不平衡的损失

不同于通用图像领域中可通过数据增强减少的数据不平衡，图表图像中不同类别之间对象数量的不平衡是自然现象，且难以消除，例如，刻度线和刻度标签的数量通常总是比图表标题和图例标签的数量多出数倍（见图3）。我们使用焦点损失 [20] 进行分类，以缓解图表图像中的对象不平衡问题：

$$L_{cls} = -\sum_{i=1}^{i=n}(i - p_i)^{\gamma}\log_b(p_i) \tag{1}$$

随后，使用平滑 L1 损失进行边界框回归，并将其与分类损失进行平衡：

$$loss_{\mu} = L_{cls}(\mu) + \lambda[\mu \geq 1]L_{loc}(t^{\mu}, v) \tag{2}$$

其中， L_{loc} 是类别 μ 的回归结果与回归目标之间的定位损失。 λ 用于调整离群值的损失，即其损失大于 1 的情况。

3.3 分类细化

数据集中图表元素的分类对检测性能至关重要。我们分析了带有图表元素标注的数据集，并理解这些元素的分类，以便对各种图表场景进行概括（见表 1）。

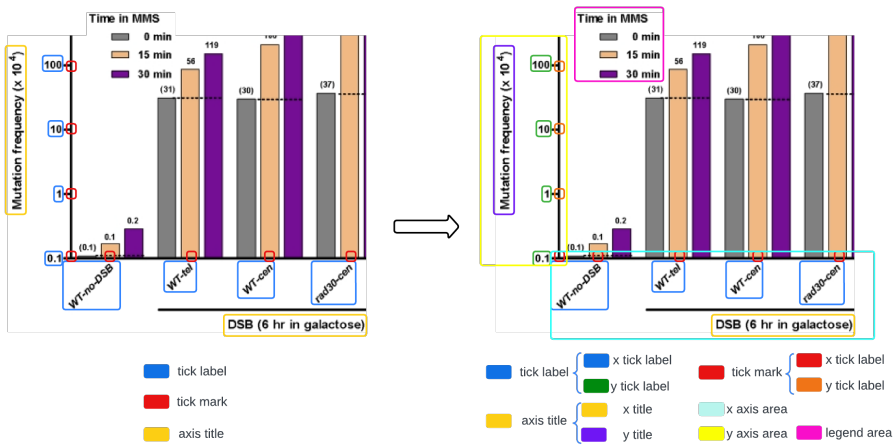


图 4：**分类细化**。图中仅展示了更新后的类别以作说明。部分元素未予显示。

在图表竞赛 [5,6], 中，Adobe Synthetic 和 PMC 数据集为图表元素检测任务提供了最有价值的标注（数据集的详细介绍见实验部分）。我们参考这两个数据集进行元素分类。表 1 的第二列列出了 Adobe Synthetic 数据集中的类别，第三列列出了 PMC 数据集中标注的类别。在 PMC 数据集中，所有与坐标轴相关的对象，如坐标轴标题、刻度标签和刻度线，均被统一标注，未区分横轴和纵轴。早期实验表明，如果将基于横轴和纵轴的项目统一视为同一类别，之后通过后处理更容易区分这些标签。这是由横轴和纵轴的定义造成的：横轴被定义为具有自变量/标签的坐标轴，而纵轴具有因变量。这类定义更多依据内容的语义，而非视觉或位置信息。因此，在使用训练后的模型进行应用时，预先按坐标轴区分这些元素有助于避免这一问题。同时，这些区分后的标签也有助于我们的方法理解图表内容，并有望获得更好的检测结果。

表 1：**分类细化**。我们研究了图表竞赛中的 Adobe Synthetic数据集和 PMC 数据集 [5,6]，并理解了第三列所示的元素类别。

	合成	PMC	细化后的类别
图表基本元素 **骨架元素**	横轴标题	坐标轴标题	横轴标题
	纵轴标题		纵轴标题
	x轴刻度标签	刻度标签	x轴刻度标签
	y轴刻度标签		y轴刻度标签
	x轴刻度线	刻度线	x轴刻度线
	y轴刻度线		y轴刻度线
	图表标题	图表标题	图表标题
	图例色块	图例标记	图例标记
	图例标签	图例标签	图例标签
	-	图例标题	图例标题
	绘图区文本	数值标签	数值标签
		标记标签	标记标签
	-	刻度分组	刻度分组
	-	其他	其他
图表 **结构区域**	绘图区	绘图区	绘图区
	-	-	横轴区域 纵轴区域 图例区域

表格 1 第二部分中的类别是另外四种结构对象标签，用于覆盖图表图像中的特定区域。我们将图表图像划分为四个最关键的结构——横轴区域、纵轴区域、绘图区和图例区域。这四个区域的定义如下：

– **绘图区**：绘图区由x轴和y轴构成。– **X/Y轴区域**：涵盖所有x/y轴刻度、x/y轴标签和x/y轴标题。– **图例区域**：涵盖所有与图例相关的项目，包括图例标签、图例标记和图例标题。

图 4 展示了一个类别细化示例。4。可以看出，每个与坐标轴相关的元素都被单独分开，而结构区域对象则覆盖图表图像中的特定区域。细化后，图表元素共归纳为 18 个类别。基于这些新类别，我们采用基于规则的方法分离与坐标轴相关的元素，并在原有绘图区的基础上额外生成三个结构区域元素，以此相应地更新 PMC 数据集。随后，为方便使用，我们将标注格式转换为标准的COCO目标检测格式。更新后的 PMC数据集及转换工具可通过第 1 节贡献摘要中的链接获取1。

4 数据集与实验

本节介绍实验中使用的现有数据集，并进行定量评估和定性分析。

4.1 数据集

Adobe Synthetic数据集。Adobe Synthetic数据集最初由 2019 [4]提出，并在 2020年图表信息图竞赛 [5]中得到改进。该数据集使用 Matplotlib 合成，包含 12 种图表类型的 14400 张图像。标注包括图表数据信息以及所有元素的标签和位置。这些标注对于图表分类、元素检测和数据提取任务具有重要价值。然而，该数据集中的图表样本多样性严重受限。ICPR 2020图表竞赛 [5] 中展示的最佳结果表明，在图表分类、文本角色分类、刻度标签关联和图例标记检测方面的准确率接近 100%，这进一步说明了数据方差有限。数据方差低通常会导致训练出的深度学习模型性能不稳定，且对数据集之外的样本泛化能力较差。我们仅参考该数据集的标注规范来细化元素分类。

PubMed Central（PMC）图表数据集。该数据集随 ICDAR 2019 [5], ICPR 2020 [6]和 ICPR 2022 [6]图表竞赛发布并更新。与 Adobe Synthetic 数据集不同，PMC 数据集是一个从 PubMed Central文献中收集并经人工标注的真实世界数据集。考虑到该数据集中的样本更多样、保真度更高，PMC 数据集成为近期大多数图表竞赛中的主要数据集。最新版本的 PMC 数据集在 ICPR 2022 CHART-Infographic竞赛 [6], 中发布，其中包含 5614 张用于图表元素检测的图像、4293 张用于最终绘图区检测和数据提取的图像，以及 22924 张用于图表分类的图像。尽管由于人工标注流程耗时，训练样本数量略显有限，但这个真实世界数据集极具价值，也比大多数合成数据集更具挑战性。

表 2： **ICPR 2022 图表竞赛中的方法比较**。 图表竞赛中任务 2 和任务 3 的评估 [6] 基于原始 PMC 分类。我们使用更新后的类别训练方法，但将预测结果回溯到原始类别，以便与图表竞赛的评估指标兼容。

团队 方法	任务 2 文本检测 任务 3 文本角色分类			
	平均交并比	召回率	精确率	F 值
六七四 _	0.435	-	-	-
IIIT_CVIT 图表理解	0.790	-	-	**0.821**
Ystar	0.810	-	-	-
UB-ChartAnalysis	0.820	-	-	0.736
Ours	**0.869**	**0.735**	**0.846**	**0.787**

Excel400K 在 [24],中，作者提出了 ExcelChart400k 数据集，其中包含总计 36 万个训练样本。该数据集使用 Excel 数据表生成，并用于训练 [24]中的绘图元素检测模型。然而，该数据集仅关注对绘图区内的数据图进行标注，大多数基本元素都未予标注。我们使用这个额外的数据集对我们方法的结果进行定性评估。

在进行初步实验后，我们观察到，加入 Adobe 合成训练数据集会降低模型在真实世界图表图像上的检测性能。这归因于合成数据集中图表渲染模式单一且样本差异有限。考虑到真实世界图表数据的高度多样性及其分布，我们使用 PMC 数据集作为训练和定量评估的主要数据集。

4.2 定量评估

为定量评估我们的方法，我们开展了三项实验：(i) ICPR 2022 图表竞赛评估 [6], ，其中我们采用了通过反向适配得到的方法结果；(ii) 使用精炼 PMC 数据集进行 COCO 目标检测评估；以及 (iii) 一项扩展实验，用于检测柱状图。

ICPR 2022 图表竞赛评估 我们将结果与 ICPR 2022 图表竞赛的结果进行比较 [6]。然而，我们的方法采用了不同的检测流程：图表竞赛将文本块检测任务视为单类别检测任务（任务 2），并将识别作为一项额外任务（任务 3）。在图表竞赛中，团队可以使用先前任务的真实标注来完成当前任务。我们的方法遵循常规检测流程（定位 + 识别）。在细化分类后，类别总数（见表 1）从竞赛中使用的原始 PMC 数据集扩展而来。考虑到这两点，我们将预测结果转换为与评估兼容的格式，并将预测结果拆分为任务 2 和任务 3，以此进行比较。表 2 展示了 2022 年官方图表竞赛的结果。我们的方法在任务 2 上优于表现最佳的“ UB-ChartAnalysis”。在

表3：**精炼 PMC 数据集上的 COCO 评估**。前4行展示了我们训练的公开检测器模型，最后3行是我们方法的消融研究。

模型	**上下文模块**	**骨干网络**	AP	AP_{50}	AP_{75}	AP_S	AP_M	AP_L
DETR	-	-	0.536	0.762	0.572	0.384	0.555	0.809
Faster R-CNN	-	ResNeXt101-FPN	0.665	0.815	0.737	0.557	0.697	0.827
Cascade R-CNN	-	ResNeXt101-FPN	0.696	0.825	0.759	0.579	0.725	0.899
Cascade R-CNN	-	SwinT-FPN	0.699	0.838	0.772	0.589	0.732	0.885
Cascade R-CNN	**PCE**	SwinT-FPN	0.708	0.842	0.775	0.591	0.742	0.903
Cascade R-CNN	**视觉上下文增强+PCE**	SwinT-FPN	0.713	**0.851**	0.786	0.597	0.741	0.909
Cascade R-CNN^L	**视觉上下文增强+PCE**	SwinT-FPN	**0.729**	0.845	**0.790**	**0.602**	**0.763**	**0.939**

任务 3 中，我们的方法无需将元素定位的真实标注作为先验知识，即可一次性检测并识别所有元素；由于检测并不完美，这可能导致任务 3 的召回率较低。总体而言，我们在任务 2 上取得了最佳检测结果，在任务 3 上取得了相当的 F 值。

在精炼 PMC 数据集上的 COCO 评估 为了更全面地概述我们方法的性能，我们使用 COCO 目标检测评估指 [21] 标，通过多种常用的两阶段检测器（见表 3）评估我们的方法。据我们所知，目前尚无在 PMC 数据集上训练的图表元素检测模型可供公开使用，以进行推理和微调。我们自行在精炼 PMC 数据集上训练了这些两阶段检测器。在表3中，第一部分展示了热门两阶段检测器的结果。由于数据集规模和 Transformer 架构，DETR 的训练收敛较慢。DETR 的小目标检测准确率也较低。图表图像中数量庞大的小元素导致整体准确率受限。在标准 Cascade R-CNN模型库中，采用 Swin-Transformer 的 Cascade R-CNN 表现最佳。第二部分展示了我们的方法在局部-全局上下文模块上的多种消融设置，该模块由视觉上下文增强（VCE）和位置上下文编码（PCE）组成。由于 Swin-Transformer 主干网络能够关注空间信息，并有望获得具备上下文感知能力的视觉特征表示，因此我们进行了消融实验，仅将位置上下文编码（PCE）集成到基于 Swin-Transformer 的 Cascade R-CNN 中。与预期一致，加入 PCE 的基于 Swin-Transformer 的 Cascade R-CNN 优于标准 Swin-Transformer Cascade R-CNN。尽管 Swin-Transformer 能够获得具备上下文感知能力的视觉特征，加入 VCE 还能进一步增强上下文提取能力，从而取得更好的结果。焦点损失与平衡平滑L1损失有助于解决样本不平衡问题，采用这种损失设置的方法达到了最先进的性能（见表 3最后一行）。

条形图检测的扩展实验虽然我们的目标是提供一种稳健的基础元素检测方法，但我们将实验扩展到了柱状图检测。我们使用 PMC条形图子集对方法进行微调，并在 PMC测试集上进行测试（见

表4： **PMC 测试数据集上的柱形图检测评估**。 对比结果来自 [25], , 我们使用相同的评估指标对我们的方法进行评估。前三列是不同交并比阈值下的 F 值，最后一列是使用图表竞赛 [5,6]评估标准计算得出的分数。

模型	交并比=0.5	IoU=0.7	IoU=0.9	分数 a_
SSD	43.65	26.28	2.67	25.83
YOLO-v3	58.84	36.14	4.14	60.97
Faster R-CNN	66.37	60.88	29.13	70.03
Faster R-CNN+特征金字塔网络	85.81	78.05	31.30	89.65
Cascade R-CNN+特征金字塔网络	86.92	83.53	55.32	91.76
我们的方法	**89.30**	**88.73**	**76.94**	**93.75**

表4）。前五个模型的结果来自 [25], , 我们采用相同的标准评估结果。前三列是采用不同交并比阈值（0.5、0.7 和 0.9）计算的 F 值。最后一列“分数 a”依据 ICPR 2022 图表竞赛 [5] 任务 6a 的评估指标计算得出。我们的方法在 PMC 数据集的条形图检测上达到了当前最先进的性能。

4.3 图表元素检测的定性评估

尽管 Excel400K 数据集没有图表基本元素的真实标注，我们仍将方法的预测结果可视化（见图5），以进行定性评估。对于 Excel400K 数据集中的大多数样本，我们的方法都能准确定位每个元素。然而，对于图5 第三行的第一个样本，我们的方法表现欠佳，因为底部附带的表格会干扰检测器，而 PMC 训练数据集中缺乏类似样本。总体而言，我们的方法能够准确定位和分类图表中的基本元素。

5 结论与未来工作

在这项工作中，我们提出了一种方法，强调图表图像中的视觉和位置上下文对于准确进行图表元素检测的重要性。我们对图表元素的类别进行了分析和精炼，以提升对各种图表设计的泛化能力，从而有助于数据解读相关的下游任务。在精炼 PMC 数据集上训练的方法，在图表元素检测任务上取得了最先进的性能。

考虑到上下文信息，文本包含丰富的信息。由于在图表图像上执行光学字符识别任务具有挑战性，许多符号很容易与字符或数字混淆，而且当文本较短时，其旋转方向也难以检测，因此我们没有将文本嵌入纳入上下文提取。未来，通过更鲁棒的光学字符识别来提取图表文本，并将图表文本嵌入作为额外的上下文信息添加到每个感兴趣区域中，或许能够提升性能。

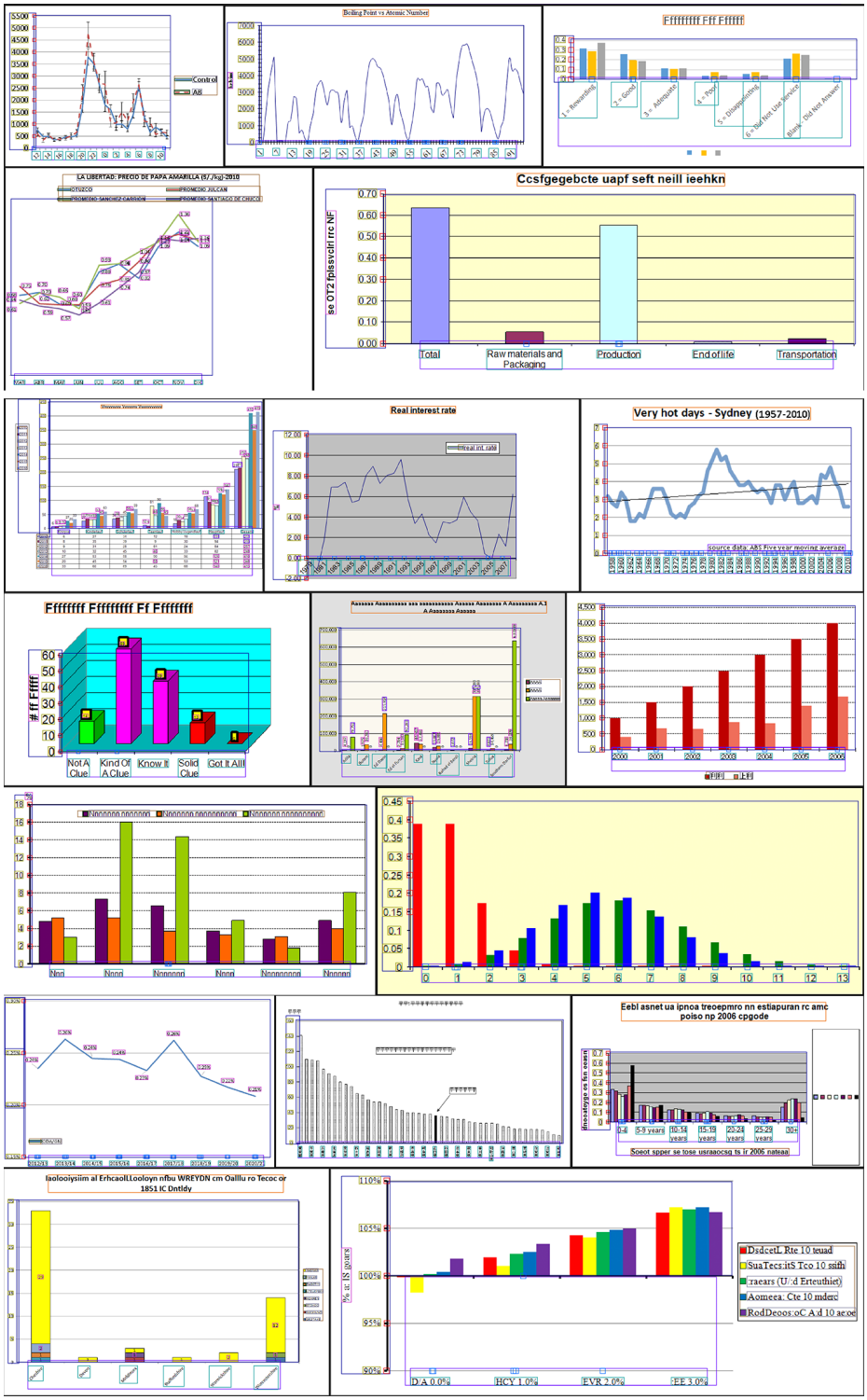


图5：我们的方法在Excel400K上的可视化。

参考文献

1. Balaji, A., Ramanathan, T., Sonathi, V.: Chart-text：一种全自动图表图像描述方法。arXiv 预印本 arXiv:1812.10636（2018） 2. Cai, Z., Vasconcelos, N.: Cascade R-CNN：深入研究高质量目标检测。载于：IEEE计算机视觉与模式识别会议论文集。第6154–6162页（2018） 3. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: 使用变换器进行端到端目标检测。载于：欧洲计算机视觉会议。第213–229页。Springer（2020） 4. Davila, K., Kota, B.U., Setlur, S., Govindaraju, V., Tensmeyer, C., Shekhar, S., Chaudhry, R.: ICDAR 2019信息图表原始表格提取竞赛（图表信息图）。载于：2019年国际文档分析与识别会议（ICDAR）。第159400203 51599 20191011092019–(). https://doi.org/./ICDAR... Davila, K., Tensmeyer, C., Shekhar, S., Singh, H., Setlur, S., Govindaraju, V.: ICPR 2020信息图表原始表格提取竞赛。载于：国际模式识别会议。第361–380页。Springer（2021） 6. Davila, K., Xu, F., Ahmed, S., Mendoza, D.A., Setlur, S., Govindaraju, V.: ICPR 2022：信息图表原始表格提取挑战赛（图表信息图）。载于：2022年第26届国际模式识别会议（ICPR）。第4995–5001页。IEEE（2022） 7. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert：用于语言理解的深度双向变换器预训练。arXiv 预印本 arXiv:1810.04805（2018） 8. Ghiasi, G., Lin, T.Y., Le, Q.V.: Nas-fpn：学习用于目标检测的可扩展特征金字塔架构。载于：IEEE/CVF计算机视觉与模式识别会议论文集。第7036–7045页（2019） 9. Girshick, R.: Fast r-cnn。载于：IEEE国际计算机视觉会议论文集。第1440–1448页（2015） 10. Girshick, R., Donahue, J., Darrell, T., Malik, J.: 用于精准目标检测和语义分割的丰富特征层次结构。载于：IEEE计算机视觉与模式识别会议论文集。第580–587页（2014） 11. Han, K., Rezende, R.S., Ham, B., Wong, K.Y.K., Cho, M., Schmid, C., Ponce, J.: Scnet：学习语义对应关系。载于：IEEE国际计算机视觉会议论文集。第1831–1840页（2017） 12. Hassan, M.Y., Singh, M., et al.: Lineex：从科学折线图中提取数据。载于：IEEE/CVF冬季计算机视觉应用会议论文集。第6213–6221页（2023） 13. He, K., Gkioxari, G., Doll´ar, P., Girshick, R.: Mask r-cnn。载于：IEEE国际计算机视觉会议论文集。第2961–2969页（2017） 14. He, K., Zhang, X., Ren, S., Sun, J.: 用于图像识别的深度残差学习。载于：IEEE计算机视觉与模式识别会议论文集。第770–778页（2016） 15. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets：面向移动视觉应用的高效卷积神经网络。arXiv 预印本 arXiv:1704.04861（2017） 16. Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K.: Squeezenet：参数量减少50倍、模型大小小于0.5 MB，精度媲美Alexnet。arXiv 预印本 arXiv:1602.07360（2016） 17. Kafle, K., Price, B., Cohen, S., Kanan, C.: Dvqa：通过问答理解数据可视化。载于：IEEE计算机视觉与模式识别会议论文集。第5648–5656页（2018）

18. Law, H., Deng, J.: Cornernet：将目标检测为成对关键点。载于：欧洲计算机视觉会议（ECCV）论文集。第734–750页（2018） 19. Li, L.H., Yatskar, M., Yin, D., Hsieh, C.J., Chang, K.W.: Visualbert：一种用于视觉与语言的简洁且高性能基线。arXiv 预印本 arXiv:1908.03557（2019） 20. Lin, T.Y., Goyal, P., Girshick, R., He, K., Doll´ar, P.: 用于密集目标检测的焦点损失。载于：IEEE国际计算机视觉会议论文集。第2980–2988页（2017） 21. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Doll´ar, P., Zitnick, C.L.: Microsoft COCO：上下文中的常见物体。载于：《计算机视觉——ECCV 2014：第13届欧洲计算机视觉会议》，苏黎世，瑞士，2014年9月6日至12日，论文集，第V部分，第13卷。第740–755页。Springer（2014） 22. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD：单次多框检测器。载于：欧洲计算机视觉会议。第21–37页。Springer（2016） 23. Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin Transformer：使用移位窗口的分层视觉变换器。arXiv 预印本 arXiv:2103.14030（2021） 24. Luo, J., Li, Z., Wang, J., Lin, C.Y.: ChartOCR：通过深度混合框架从图表图像中提取数据。载于：IEEE/CVF冬季计算机视觉应用会议论文集。第1917–1925页（2021） 25. Ma, W., Zhang, H., Yan, S., Yao, G., Huang, Y., Li, H., Wu, Y., Jin, L.: 朝着高效的图表图像数据提取框架迈进。arXiv 预印本 arXiv:2105.02039（2021） 26. Oglan, V.A.: Chart Sense：用常识图表教授3至8年级的信息文本和文学作品。《语言艺术》**92**(5)，368（2015） 27. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You Only Look Once：统一的实时目标检测。载于：IEEE计算机视觉与模式识别会议论文集。第779–788页（2016） 28. Redmon, J., Farhadi, A.: YOLO9000：更好、更快、更强。载于：IEEE计算机视觉与模式识别会议论文集。第7263–7271页（2017） 29. Redmon, J., Farhadi, A.: YOLOv3：渐进式改进。arXiv 预印本 arXiv:1804.02767（2018） 30. Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN：借助区域提议网络实现实时目标检测。神经信息处理系统进展**28**，91–99（2015） 31. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C.: MobileNetV2：倒置残差和线性瓶颈。载于：IEEE计算机视觉与模式识别会议论文集。第4510–4520页（2018） 32. Simonyan, K., Zisser man, A.: 用于大规模图像识别的超深卷积网络。arXiv 预印本 arXiv:1409.1556（2014） 33. Tan, H., Bansal, M.: LXMERT：从变换器中学习跨模态编码器表示。arXiv 预印本 arXiv:1908.07490（2019） 34. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention Is All You Need。载于：神经信息处理系统进展。第5998–6008页（2017） 35. Xie, S., Girshick, R., Doll´ar, P., Tu, Z., He, K.: 用于深度神经网络的聚合残差变换。载于：IEEE计算机视觉与模式识别会议论文集。第1492–1500页（2017） 36. Zhang, X., Zhou, X., Lin, M., Sun, J.: ShuffleNet：一种面向移动设备的极高效卷积神经网络。载于：IEEE计算机视觉与模式识别会议论文集。第6848–6856页（2018）