arXiv:2305.04833v2 [cs.IR] 8 Nov 2023

# 重新审视视觉丰富文档的表格检测数据集

# 文档

肖斌[1], Murat Simsek[1], Burak Kantarci[1*]和 Ala Abu Alkheir[21]渥太华大学电气工程与计算机科学学院，地址：加拿大安大略省渥太华，邮编 K1N 6N5。[2]Lytica, 555 Legget Dr，地址：加拿大安大略省渥太华，邮编 K2K 2X3。

通讯作者。电子邮箱：burak.kantarci@uottawa.ca；贡献作者：bxiao103@uottawa.ca；murat.simsek@uottawa.ca；ala abualkheir@lytica.com；

**摘要**

随着电子文档数量激增，表格检测已成为理解视觉丰富文档的一项基础任务。然而，相关研究中广泛使用的热门公开数据集存在固有局限，包括样本带有噪声且不一致、训练样本有限以及数据来源有限。这些局限使得这些数据集无法可靠地评估模型性能，也无法反映模型的实际能力。因此，本研究重新审视了一些具有高质量标注的开放数据集，识别并清理其中的噪声，并统一这些数据集的标注定义，以合并成一个更大的数据集，称为 Open-Tables。此外，为丰富数据来源，我们使用信息与通信技术（ICT）商品的 PDF 文件构建了一个新的 ICT-TD 数据集。该数据集来自不同领域，包含在开放数据集中鲜少出现的独特样本。为确保数据集的标签质量，我们遵循领域专家的指导，手动标注了该数据集。所提出的数据集具有挑战性，可作为商业场景中实际案例的样本。我们使用多种先进的目标检测模型构建了强大的基线。实验结果表明，尽管现有开放数据集的数据来源不同，但它们之间的领域差异很小。所提出的 Open-Tables 和 ICT-TD 数据集具有高质量且一致的标注，因此能够为模型提供更可靠的评估。此外，它们也更适用于跨域设置。实验结果表明，在跨域设置下，使用清理后的 Open-Tables 数据集训练的基准模型，其加权平均F1值比使用带噪声版本 Open-Tables 训练的相应模型高出 0.6%–2.6%，这证明了所提出数据集的可靠性。这些数据集可从以下网址公开获取：http://ieee-dataport.org/documents/table-detection-dataset-visually-rich-documents。

**关键词：**目标检测、表格检测数据集、ICT供应链、表格检测

# 1 引言

表格或表格数据已广泛应用于电子文档中，用于概括关键信息，以便向人类读者高效地呈现这些信息。然而，电子文档（例如便携式文档格式（PDF）文件）无法提供足够的元数据来描述这些表格的位置和结构，这意味着这些表格缺乏结构化信息，无法被自动快速处理和解析。随着电子文档数量的激增，表格检测（TD）已成为下游文档理解任务中的一项基础任务，例如关键信息提取和表格结构识别 [1]。随着深度学习的发展，将电子文档转换为视觉信息丰富的文档图像，并将该问题建模为目标检测问题，已成为主流解决方案。目前，已有一些针对表格检测问题的公开数据集，例如 ICDAR2013 [2], ICDAR2017 [3], ICDAR2019 [4]和 TableBank [5]。其中一些数据集采用手动标注，这意味着其标注更加可靠且一致，但这些数据集中的训练样本数量通常有限。此外，这些数据集采用的标注定义往往各不相同，因此我们无法简单地将它们合并，以构建规模更大的数据集。相比之下，TableBank [5] 和 PubLayNet [6] 通过解析电子文档的元数据进行标注，因此这些标注存在噪声且不一致，尽管这些数据集规模要大得多。图1 展示了来自 Table-Bank 测试集的两个样本。此类由元数据生成的数据集的一个典型问题是，边界框可能比理想边界框更大，如图 1 (a) 所示；当交并比（IoU）阈值较高时，这会导致评估结果不可靠。另一个问题是，有些表格缺失，或者其边界框不够大，无法覆盖整张表格，如图1(b) 所示。表格检测数据集的质量对表格检测问题至关重要，因为成功的表格检测应用应避免丢失表格中呈现的信息。测试集中的噪声标签会影响模型评估，尤其是在使用由交并比分数设定阈值的常用评估指标时。值得一提的是，尽管人工标注数据集的标注质量更高，但其训练集和测试集中仍存在许多噪声样本。因此，本研究重新审视了若干标注质量良好的数据集，包括 ICDAR2013、ICDAR2017、ICDAR2019、Marmot 和 TNCR，统一这些数据集的标注定义，清理噪声样本，并将其合并为一个更大的数据集，称为 Open-Tables。新的 Open-Tables 数据集能够最大限度地减少噪声样本对模型评估的副作用，并提供更可靠的结果。我们还收录了

关于 Open-Tables 数据集的更多详情，请参见第3.1节。

除了噪声标签的问题外，开放数据集的数据来源也十分有限，主要来自学术出版物或公开的政府文档。由于这些文档必须遵循一系列写作规范，其数据来源有限，导致文档内部和类别间的差异较小。换言之，检测模型在这些数据集上很容易取得出色的性能，但这些数据集也无法体现真实企业应用的复杂性。因此，本研究使用信息与通信技术（ICT）领域的数据表，提出了一个新的表格检测数据集。由于样本、布局和外观具有领域特有的差异，该数据集更具挑战性。图2 展示了所提出的数据集中的一些样本，这些样本在公开数据集中几乎找不到。例如，图2 (a) 是一个包含多个子表格的表格，而图2 (7) 是一个将图形作为部分单元格内容的表格。图2列出了所提出的数据集中的几个示例，这使其与其他开放数据集有所不同，也更具挑战性。我们将在第3.2节进一步介绍所提出的数据集。

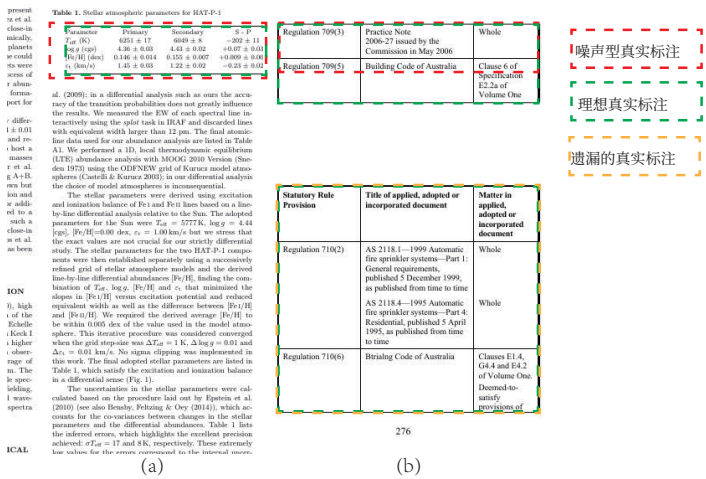
我们还使用先进的目标检测模型构建了强大的基线，包括用于 Open-Tables 和 ICT-TD 数据集的 TableDet [7]，DiffusionDet [8]，可变形 DETR [9] 和 SparseR-CNN [10] 。由于 Open-Tables 与 ICT-TD 数据集之间存在明显的领域差异，我们还针对跨领域设置构建了基线，以进一步评估模型的泛化能力。Open-Tables 和 ICT-TD 数据集能够推动对典型及跨领域表格检测问题的研究，并最大限度地减少噪声样本对模型评估的负面影响。

总而言之，本文的贡献主要体现在以下三个方面：

1. 本研究重新审视了几个开放的表格检测数据集，统一其标注，清理噪声型标签，并将它们合并为一个名为 Open-Tables 的新数据集。Open-Tables 数据集能够提供更可靠的模型评估，同时最大限度地减少噪声型标签的副作用。

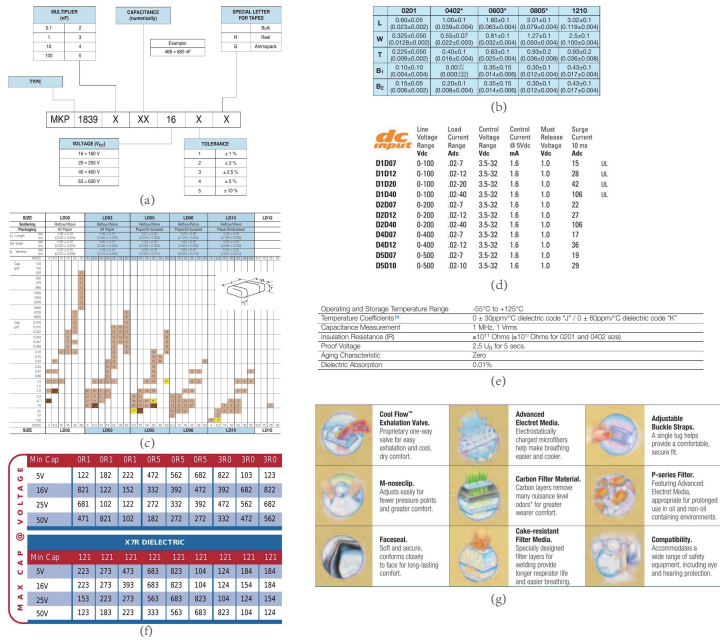
2. 本文提出了一个新的手动标注数据集 ICT-TD，该数据集使用包含大量领域特定训练样本的 ICT 商品 PDF文件构建。ICT-TD 提供了新的数据来源，并包含许多独特样本，这些样本很难在其他开放数据集中找到。

3. 除了采用不同类型先进目标




(a) (b)

**图1**：来自 TableBank 的两个噪声样本 [5] 测试集。红色边界框是数据集提供的真实标注框。在图(a)中，边界框大于理想边界框，因此在 IoU 阈值较高时，评估结果不可靠。图(b)展示了一个样本，其边界框需要扩大才能覆盖整个表格。此外，图(b)底部的表格未被标注。值得一提的是，图像分辨率较低是由数据集提供的图像所致。



(a) (b) (c) (d) (e) (f) (g)

**图2**：所提出的 ICT-TD 数据集中的表格示例。

检测方法的多种强基线外，本研究还提供了跨领域设置下的基准。

本文其余部分的组织如下：第2节讨论相关研究，包括相关数据集以及目标检测模型方面的研究

和表格检测模型。第3 节介绍我们所提出的 Open-Tables 和 ICT-TD 数据集、形式化问题定义以及基线模型。第4 节介绍实验设置和结果，并为我们所提出的数据集构建基线。最后，我们在第5节给出结论并讨论可能的研究方向。

## 2 相关工作

如第 1 节所述，表格检测问题通常被构建为目标检测问题。本节将讨论相关数据集、热门目标检测模型和表格检测模型。

### 2.1 相关数据集

已有许多可用于表格检测问题的公开数据集。其中一些数据集专为表格检测问题创建，例如 ICDAR2013 [2], ICDAR2017 [3], ICDAR2019 [4] 和 TNCR [11]。ICDAR2013 数据集是许多研究中广泛使用的基准，其中包含从公共政府文档中收集的 238 张图像。ICDAR2013 数据集中的一些图像不包含任何表格，因此共有 150 个表格。由于该数据集规模相对较小，许多研究按照 ICDAR2013 竞赛的评估指标设置 [2]，在交并比阈值为 0.5 时取得了 100% 的 F1分数。ICDAR2017 和 ICDAR2019 数据集是相应竞赛中使用的另外两个数据集。ICDAR2017 数据集由包含 1600 张图像的训练集和包含 817 张图像的测试集组成。ICDAR2017 中的文档收集自 CiteSeer，这是一个学术出版物来源。ICDAR2019 包含两个独立的集合，分别为档案文档图像集合和现代文档图像集合。在本研究中，我们仅考虑 ICDAR2019 数据集中的现代文档集合，其中包含 600 张训练图像和 240 张测试图像。TNCR [11] 同样收集公共政府文档，并进一步定义了五种表格类型，包括全线表格、无线表格、合并单元格表格、部分有线表格和部分有线合并单元格表格。TNCR 的训练、测试和验证图像分别为 4634 张、987 张和 1000 张。IIIT-AR-13K [12]是另一个数据集，定义了五种文档对象类型，包括表格、图形、自然图像、徽标和签名。IIIT-AR-13K 数据集以商业年度报告为数据来源，包含 9000 张训练图像、2000 张测试图像和 2000 张验证图像。

一些数据集是通过解析 Microsoft Word 文件或 LaTeX 源文件的元数据生成的，通常比人工标注数据集大得多。TableBank [5] 是一个典型的自动生成数据集，包含由 LaTeX 和 Microsoft Word 生成的样本。Word 文件从互联网上收集，而 LaTeX 源文件

代码来自开放的学术出版物。同样，PubLayNet [6]数据集是通过解析学术出版物的 XML 格式文档生成的。PubLayNet 定义了五种文档组件，包括文本、标题、列表、表格和图形。我们在表格1中总结了这些开放数据集的统计信息。

如第1节所述，开放数据集存在一些固有局限。首先，这些数据集的标注定义可能有所不同，尤其是对于歧义样本。其次，这些数据集的数据来源有限，主要来自学术出版物和开放的政府文件。尽管还有其他数据来源，例如商业年度报告，但这些来源中的样本通常较为简单，且与其他开放数据集中的样本相似。第三，这些开放数据集可能含有噪声，尤其是自动生成的数据集，这会导致使用这些数据集进行的模型评估不可靠。此外，这些数据集的数据来源非常有限，主要来自学术出版物和开放的政府文件。这些局限使得使用这些数据集训练的模型难以对不同领域实现良好的泛化能力。此外，训练集中的噪声样本会影响模型性能，而测试集中的噪声样本则会使模型评估变得不可靠。因此，为了缓解这些局限，本研究提出了 Open-Tables 和 ICT-TD 数据集，它们具有一致的标注且噪声较少。此外，本研究还为跨域设置构建了基准，这一设置对表格检测应用而言颇具挑战，但也很有价值。

## 2.2 目标检测模型

目标检测是计算机视觉中的热门问题，近年来得到了广泛讨论。我们可以将基于深度学习的目标检测模型分为三类：单阶段、两阶段和基于变换器的模型。单阶段和两阶段模型都高度依赖卷积神经网络（CNN），并采用区域检测与目标分类的流程。这两类模型的主要区别在于，这两个任务能否在单个深度神经网络中完成。以 YOLO [13] 为单阶段模型的例子，YOLO 的第一步是将输入图像划分为 N * N 个网格单元。每个网格单元用于预测其中包含目标对象的置信度分数，以及目标类别的条件概率。所有这些步骤都在单个模型中完成。相比之下，两阶段模型（如 Faster-RCNN [14], ）通常会使用区域提议网络（RPN）生成一系列区域提议。这些区域提议会被输入 ROIHead，以对目标类别进行分类，并通过回归任务细化边界框。在两阶段模型中，区域提议由独立网络生成，因此这类模型有两个

阶段。就推理时间而言，单阶段模型通常比两阶段模型更快；而在检测性能方面，两阶段模型通常优于单阶段模型。已有许多研究[15, 13, 16, 17, 18, 19, 20, 10, 21]遵循这些典型的单阶段和两阶段模型设计，对模型的某些部分进行改进，例如区域提议方法、骨干网络，或在多任务架构中添加其他子任务。随着自注意力机制的发展[22],变换器架构也被应用于目标检测问题。DETR [9]是首项将变换器架构与目标检测问题相结合的研究。在 DETR 中，嵌入网络采用一种广泛使用的 CNN 网络来生成图像特征，随后将这些特征输入由变换器实现的编码器—解码器架构，并将该问题表述为二分图匹配问题。DETR 能够达到最先进的性能，但收敛速度非常慢。因此，一些研究（如 Deformable DETR） [9], Conditional DETR [23], 动态 DETR [24]和 Fast DETR [25],尝试通过引入不同的位置编码方法并改进注意力模块来提升性能、加快 DETR 的速度。扩散模型广泛用于图像生成问题，并最初应用于目标检测问题的 DiffusionDet [8]。与典型的扩散模型类似，DiffusionDet [8] 也包含正向和反向扩散过程。不过，这两个过程的目标都是边界框，即对其进行加噪和去噪。

## 2.3 表格检测

表格检测是关键信息提取和视觉丰富型文档理解等下游任务中的基础步骤。通常，表格检测问题被定义为目标检测问题。考虑到自然图像中的物体与文档图像中的表格之间存在差异，YOLOv3-TD [26] 提出了一种锚框优化策略和两种后处理方法，以调整检测方法。YOLOv3-TD 的作者观察到，除非表格非常大，否则表格的宽度通常大于其高度。基于这一观察，他们提出了一种基于 K-means 的方法来优化锚框，从而获得更多“水平型”锚框。此外，他们还通过从预测区域中去除白色空白边距，并过滤噪声型页面对象，将这两种操作作为后处理方法，以进一步提升模型性能。CascadeTabNet [27]是一种应用于表格检测问题的典型两阶段模型。CascadeTabNet 基于 Cascade Mask R-CNN [28]，并采用 HRNet [29] 作为骨干网络。此外，CascadeTabNet 还采用了一种

图像增强方法，可以加粗文本区域并减少空白区域，并采用迁移学习方法迭代训练模型。与单阶段和两阶段模型类似，一些研究也讨论了用于 [30] 表格检测问题的基于Transformer的方法，例如 DETR [31]。还有许多其他研究讨论了表格检测问题，例如 TableDet [7], Decnt [32], Deep-DeSRT [33], Tablenet [34], ，其中大多数研究遵循目标检测范式，并使用上述不同类型的目标检测模型。值得注意的是，研究 [35, 36] 也讨论了 ICT 领域的表格检测问题，但这两项研究使用的数据集都小于本研究提出的 ICT-TD 数据集。</style}

# 3 所提出的数据集

如前所述，现有的开源表格检测数据集通常不够复杂，难以反映商业场景的复杂性和难度。本节将详细介绍所提出的 Open-Tables 和 ICT-TD 数据集。

## 3.1 Open-Tables 数据集

在本节中，我们讨论 ICDAR2013 [2],ICDAR 2017 [3],ICDAR2019 [4], Marmot [37], 的噪声清理和标注对齐，以及 TNCR [11]数据集以创建 Open-Tables 数据集。

如第2.1节所述，ICDAR2019 包含档案文档和现代文档。本研究仅使用现代文档。由于 TNCR 数据集中有五种细粒度的表格类型，我们将所有这些标注统一转换为一种类型，即表格。尽管此处使用的数据集标注质量相对较高，但许多样本仍存在噪声标注，其问题如图1所示。测试集中的这些样本会影响模型评估，而训练集中的噪声样本则会降低模型性能。因此，我们首先修正了这些噪声样本。

除了噪声型标注外，正如第 1 节和 2.1 节所述，这些开放数据集对表格的定义也可能存在差异。这个问题由歧义样本导致。图 3展示了 TNCR 数据集中的两个歧义样本。第一个歧义样本中有两个备选边界框，可能导致标注不一致。TNCR 数据集提供的真实标注（绿色框）不包含表格说明部分。第二个歧义样本被标注为表格，但在其他数据集中没有必要这样标注，因为它也可以被定义为文档页脚。为处理这些歧义样本，我们制定了以下规则来统一数据集。

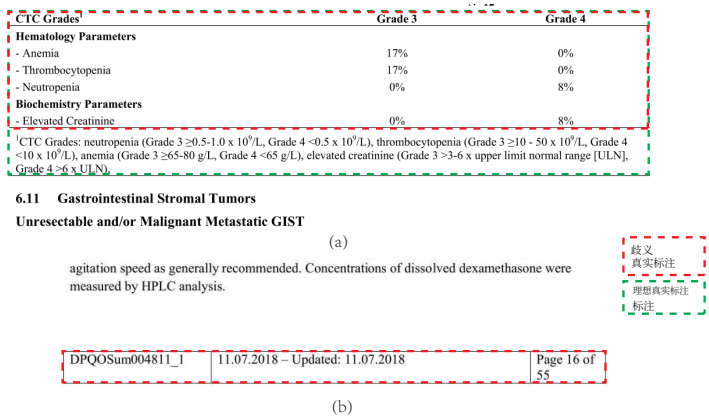
首先，我们优先考虑表格线，也就是说，纳入表格线所围住的所有内容。然而，如果表格线没有围住表格说明部分，就不应将其定义为表格的一部分。其次，表格至少应有两条线和两列。按照这两条规则，我们应将第一个样本的表格说明部分纳入表格，并将第二个样本定义为非表格，如图3所示。

## 3.2 ICT-TD 数据集

本节将介绍我们提出的 ICT-TD 数据集的数据收集和数据预处理。我们收集了 175,682 份 PDF 文档，涵盖 370 种不同的 ICT 商品。由于每个 PDF 文件可能包含多页，我们将每页转换为分辨率为每英寸 200 点的图像，共得到 3,581,805 张图像。我们采用随机抽样方法，从这些图像中选出 5,000 个包含表格的样本，并手动标注图像中所有表格的边界框。为便于比较，我们在表 1中总结了 ICT-TD 数据集和一些公开数据集的统计信息。ICDAR2013 [2]是一个小型数据集，且未提供训练集。ICDAR2017 [3], ICDAR2019 [4], Marmot [37],和 TableBank [5] 均使用学术出版物或公共政府文档作为数据来源，因此这些数据集无法反映真实企业场景的复杂性，也难以适用于 ICT 领域。

由于 ICT 领域的文档中存在许多模糊案例，我们制定了以下表格标注规则。首先，表格至少应包含两行和两列，因为表格应当是关键信息的汇总。只有一行或一列的内容，我们视为纯文本或图形。图 4展示了一个单行图示例，其外观类似表格，但根据这条规则不应标注为表格，而应视为图形。其次，表格应包含描述商品的信息，因为我们希望为特定领域的应用提取信息。某些信息的格式可能类似表格，例如文档的索引页，但对下游任务并无用处。图 5展示了两个未被标注为表格的样本，因为它们是内容索引，不包含任何商品的信息。然而，它们的外观与表格相似，这使该数据集更具挑战性。第三，除非有线条将表格标题和表格注释划分为表格的一部分，否则不应将它们纳入表格，因为本研究聚焦于表格检测问题。表格标题和表格注释应视为文档中的不同组成部分，这超出了本研究的范围。

根据这些规则，所提出的 ICT-TD 数据集中的表格可根据内容和结构分为四类



| CTC Grades[1] | Grade 3 | Grade 4 |
|---|---|---|
| **Hematology Parameters** | | |
| - Anemia | 17% | 0% |
| - Thrombocytopenia | 17% | 0% |
| - Neutropenia | 0% | 8% |
| **Biochemistry Parameters** | | |
| - Elevated Creatinine | 0% | 8% |

[1]CTC Grades: neutropenia (Grade 3 ≥0.5-1.0 x $10^9$/L, Grade 4 <0.5 x $10^9$/L), thrombocytopenia (Grade 3 ≥10 - 50 x $10^9$/L, Grade 4 <10 x $10^9$/L), anemia (Grade 3 ≥65-80 g/L, Grade 4 <65 g/L), elevated creatinine (Grade 3 >3-6 x upper limit normal range [ULN], Grade 4 >6 x ULN),

**6.11 Gastrointestinal Stromal Tumors**

**Unresectable and/or Malignant Metastatic GIST**

(a)

agitation speed as generally recommended. Concentrations of dissolved dexamethasone were measured by HPLC analysis.

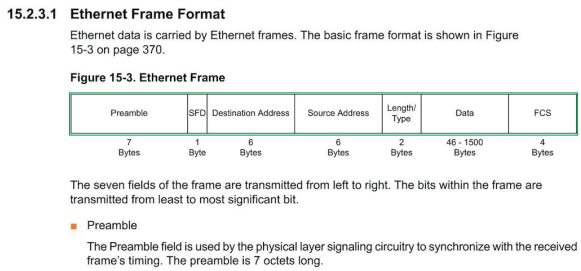
歧义真实标注

理想真实标注标注

| DPQOSum004811_1 | 11.07.2018 – Updated: 11.07.2018 | Page 16 of 55 |
|---|---|---|

(b)

**图 3**：两个歧义样本。图(a)展示了表格的两种替代标注。图(a)中的绿色边界框是 TNCR 数据集提供的真实标注，其中不包含表格说明部分。图(b)中的红色边界框被定义为真实标注，但在其他数据集中不属于表格。



**15.2.3.1 Ethernet Frame Format**

Ethernet data is carried by Ethernet frames. The basic frame format is shown in Figure 15-3 on page 370.

**Figure 15-3. Ethernet Frame**

| Preamble | SFD | Destination Address | Source Address | Length/ Type | Data | FCS |
|---|---|---|---|---|---|---|
| 7 Bytes | 1 Byte | 6 Bytes | 6 Bytes | 2 Bytes | 46 - 1500 Bytes | 4 Bytes |

The seven fields of the frame are transmitted from left to right. The bits within the frame are transmitted from least to most significant bit.

- Preamble

  The Preamble field is used by the physical layer signaling circuitry to synchronize with the received frame's timing. The preamble is 7 octets long.

**图 4**：一个未被标注为表格的单行图示例。我们用绿色方框标出该单行图。

：全线表格、部分线框表格、无线框表格和其他特殊表格。图 2展示了所提出的 ICT-TD 数据集中这些不同类型表格的一些样本。图2(a) 是一个由许多子表格组成的特殊表格。由于每个子表格都描述了商品的一个参数，我们将这些子表格的并集视为一个单独的特殊表格。

# 4 实验结果与分析

## 4.1 主要结果

在本节中，我们开展实验，为所提出的数据集建立基线。我们选择了四种最先进的方法作为基线模型，包括 TableDet [7]，、DiffusionDet [8]，、Deformable-DETR [9]和 SparseR-CNN [10]。TableDet 基于 Cascade-RCNN [38]构建，并利用迁移学习和表格感知数据增强，进一步提升表格检测问题上的性能。Deformable-DETR 是一种典型的基于 Transformer 的方法，而 DiffusionDet 将扩散过程 [39, 40] 引入目标检测问题，并采用随机区域提议。SparseR-CNN 是一种采用可学习区域提议的典型方法。因此，我们的基线模型包括一个两阶段模型、一个基于 Transformer 的

**表 1**：ICT-TD 数据集和公开数据集的统计信息。值得注意的是，此表中的数值表示图像数量，而非表格数量。* 表示这些数据集用于创建 Open-Tables 数据集。

| 数据集 | Type | 训练 | 测试 | 验证集 | 数据来源 |
|---|---|---|---|---|---|
| TableBank [5] | 生成 | 130，463 | 5，000 | 10，000 | Word 和 LaTeX |
| PubLayNet [6] | 生成 | 86，950 | 3，772 | 3，950 | PubMed |
| ICDAR2013* [2] | 手动 | – | 150 | – | 政府文件 |
| ICDAR2017* [3] | 手动 | 1600 | 817 | – | CiteSeer |
| ICDAR2019* [4] | 手动 | 600 | 240 | – | 期刊、表格、财务报表 |
| Marmot* [37] | 手动 | 2000 | – | – | 电子书和 CiteSeer |
| TNCR* [11] | 手动 | 4634 | 987 | 1000 | 政府文件 |
| IIIT-AR-13K [12] | 手动 | 9000 | 2000 | 2000 | 年度报告 |
| Open-Tables | 手动 | 8834 | 1240 | 1000 | 合并数据集 |
| ICT-TD | 手动 | 4000 | 1000 | – | ICT PDF 文档 |

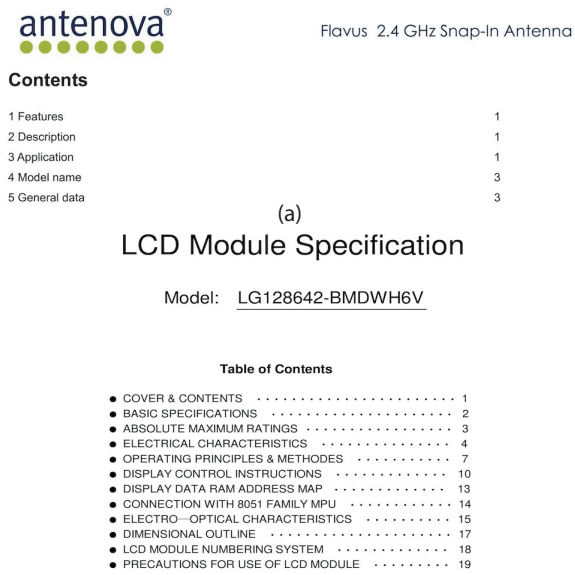



**图 5**：由于这两个样本描述的并非 ICT 商品，且对下游任务没有帮助，因此未将其标注为表格。图(a)展示了一个外观类似于无框线表格的样本，但根据我们定义的标注规则，该样本未被标注为表格。图(b)展示了一个被描述为表格的样本，但根据我们定义的标注规则，该样本未被标注为表格。

模型、一个采用随机提议的模型，以及一个采用可学习区域提议的模型，以涵盖最常见的目标检测器。值得需要说明的是，我们没有纳入单阶段检测器，因为单阶段检测器通常不如本文纳入的模型。我们使用 Detectron 重新实现了 TableDet2 [41]，同时保留了表格感知增强方法。Deformable-DETR 的实现可在 detrex 中找到 [42]。DiffusionDet 和 SparseR-CNN 均有官方实现。所有这些基线模型都使用在 ImageNet 50 [43], 上预训练的 ResNet [44] 作为训练的起点。值得注意的是，TableDet 的原始设计使用在 COCO 数据集上预训练的 Cascade MaskR-CNN [45]进行初始化。我们遵循这些基准模型默认的模型参数配置，但针对 DiffusionDet、Deformable-DETR 和 SparseR-CNN 的训练调度调整了部分参数，因为它们默认的训练调度参数是基于 COCO 数据集调优的。我们在表2中总结了这些调度参数。值得一提的是

所有这些基准模型均基于 Detectron2 构建。因此，我们遵循 Detectron2 在表格 2中的术语。

我们使用精确率、召回率和 F1分数作为评估指标。我们使用 IoU 分数作为阈值，以判断表格是否被检测到，该分数可通过公式1计算。随后，真正例是指其与某个真实标注边界框的 IoU 分数大于 IoU 阈值的预测数量，这些对应的真实标注边界框则视为已检测到。同理，假正例是指其与所有真实标注边界框的 IoU 均小于 IoU 阈值的预测数量；假负例则是指未被检测到的真实标注边界框数量。最后，精确率、召回率和 F1分数分别可通过公式 2、3 和 4计算。

如第 1 节所述，表格检测问题要求检测器在交并比阈值较高时仍保持足够的精确率和召回率，而较高交并比阈值下的分数更具区分性。因此，我们遵循 ICDAR2019 竞赛的做法 [4] ，将加权 F1分数作为主要评估指标，其定义见公式 5。我们选择 80%、85%、90% 和 95% 作为交并比阈值，而不是 ICDAR2019 竞赛中采用的 60%、70%、80% 和 90%。此外，我们还遵循了研究中的实验设置中 [11], 提供了不同交并比阈值（从 50% 到 95%）下精确率、召回率和 F1分数的详细实验结果。我们在 6.1 节中列出这些详细实验结果。

图形6 和7 展示了基线模型的一些预测样本。子图 (a)、(b)、(c)、(d) 分别是 TableDet、DiffusionDet、Deformable-DETR 和 SparseR-CNN 的结果。对于图6中的表格，其真实标注应包含表格说明部分，因为底部的一条线框住了说明文本。然而，TableDet 的预测框不够大，无法覆盖所有说明文本。Deformable-DETR 没有将说明文本视为表格的一部分，但其预测框能够很好地贴合其他部分。相比之下，DiffusionDet 和 SparseR-CNN 能够很好地检测出该表格。对于图7中的表格，TableDet 和 DiffusionDet 能够成功检测出两个表格，尽管它们的预测框无法精确贴合表格。相比之下，Deformable-DETR 将两个表格检测为一个，而 SparseR-CNN 漏检了底部的第二个表格。这些样本表明，基线模型在检测所提出的 Open-Tables 数据集中的表格时各有不足，说明 Open-Tables 数据集可作为表格检测研究的有用资源。同样，我们在图8 和9中展示了 ICT-TD 数据集上的若干预测样本。图8中表格的理想框应覆盖表头单元格，而这些单元格

**表格 2**：基准模型的关键参数。

| 方法 | TableDet | DiffusionDet | Deformable-DETR | SparseR-CNN |
|---|---|---|---|---|
| 优化器 | SGD | AdamW | AdamW | AdamW |
| 最大迭代次数 | 25,000 | 50,000 | 50,000 | 50,000 |
| 最大训练轮数 | 100 | 200 | 200 | 200 |
| 步骤 | - | 37,500 | 37,500 | 37,500 |
| 调度器 | - | 多步学习率调度器 | 多步学习率调度器 | 多步学习率调度器 |
| 基础学习率 | 1.0e-03 | 1.0e-05 | 1.0e-04 | 2.5e-05 |
| 伽马值 | - | 0.1 | 0.1 | 0.1 |
| 每批图像数 | 16 | 16 | 16 | 16 |

没有被线条框住。然而，只有 TabeDet 能覆盖这些表头单元格，但它将上方的表格检测成了两个表格。图9 展示了一个样本，其中四种基线模型都将一幅图形识别为表格。值得一提的是，图8 和9 中的样本具有领域特异性，因此 ICT-TD 数据集可作为 ICT 领域及跨领域应用的有用资源。

$$IoU = \frac{Overlap\ Area\ of\ two\ Boxes}{Union\ Area\ of\ two\ Boxes} \tag{1}$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}} \tag{2}$$

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}} \tag{3}$$

$$\text{F1-score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{Weighted Avg. F1-score} = \frac{\sum_{i=1}^{4} IoU_i \cdot F1@IoU_i}{\sum_{i=1}^{4} IoU_i} \tag{5}$$

## 4.2 跨领域表格检测

在本节中，我们讨论在跨域设置下使用该提议数据集的潜力。正如第3.2节所讨论的，ICDAR2013、ICDAR2017、ICDAR2019、Marmot 和 TNCR 也是具有高质量标注的人工标注数据集，其数据来源包括学术出版物和公开的政府文件。因此，我们将这些数据集合并，得到一个名为 Open-Tables 的新数据集，其中包含 8834 个训练样本、2282 个测试样本和 1000 个验证样本。值得一提的是，正如第 2节所讨论的，TNCR 包含五种不同类型的表格组。我们简单地将所有这些组别合并为一个表格组。在完成第 3.1节所讨论的数据清理任务后，我们采用两种跨域设置来构建基准。首先，我们使用 ICT-TD 的训练集进行训练

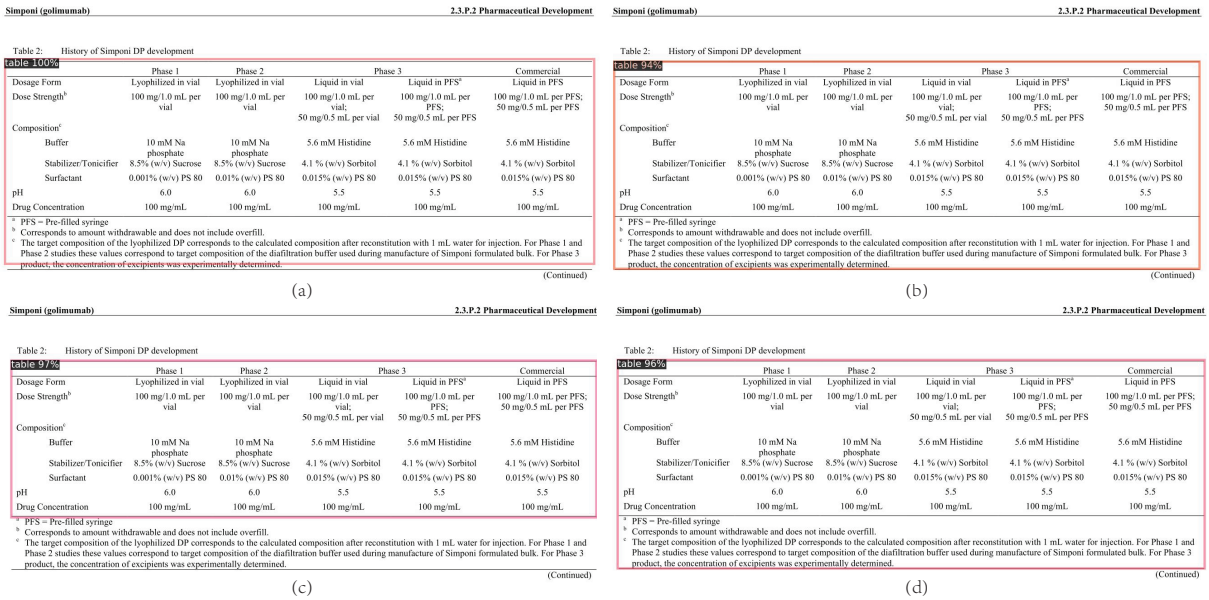



**图6**：基线模型在 Open-Tables 测试集上的预测样本。图形 (a) (b) (c) (d) 分别为 TableDet、DiffusionDet、Deformable-DETR 和 SparseR-CNN 的结果。子图中的置信度分数分别为 100%、94%、97% 和 96%。

**表3**：ICT-TD 数据集上的 F1分数实验结果。4000 和 1000 是样本数量。

| 数据集 | | 模型 | 不同 IoU阈值下的 F1 分数 | | | | 加权平均F1 |
|---|---|---|---|---|---|---|---|
| 训练 | 测试 | | 80% | 85% | 90% | 95% | |
| ICT-TD 训练集 (4000) | ICT-TD 测试集 (1000) | TableDet | 93.9 | 92.4 | 89.6 | 75.9 | 87.5 |
| | | DiffusionDet | 95.5 | 94.2 | 91.3 | 76.5 | 88.9 |
| | | Deformable-DETR | 95.1 | 93.8 | 91.6 | 82.1 | 90.3 |
| | | SparseR-CNN | 94.3 | 92.9 | 90.4 | 79.3 | 88.9 |

用于评估模型性能的检测基线模型和 Open-Tables 测试集。该设置下的实验结果如表5所示。在第二种设置中，我们使用 Open-Tables 数据集的训练集和 ICT-TD 数据集的测试集构建基准，结果如表6所示。值得一提的是，我们使用与第 4.1 节相同的评估指标，表6中的 * 表示模型使用噪声型 Open-Tables 数据集的版本进行训练。

实验结果表明，在 ICT-TD 数据集上表现最佳的 Deformable-DETR，在跨域设置下也具有最强的泛化能力。然而，跨域设置的挑战性要大得多，所有基准模型的性能都.....

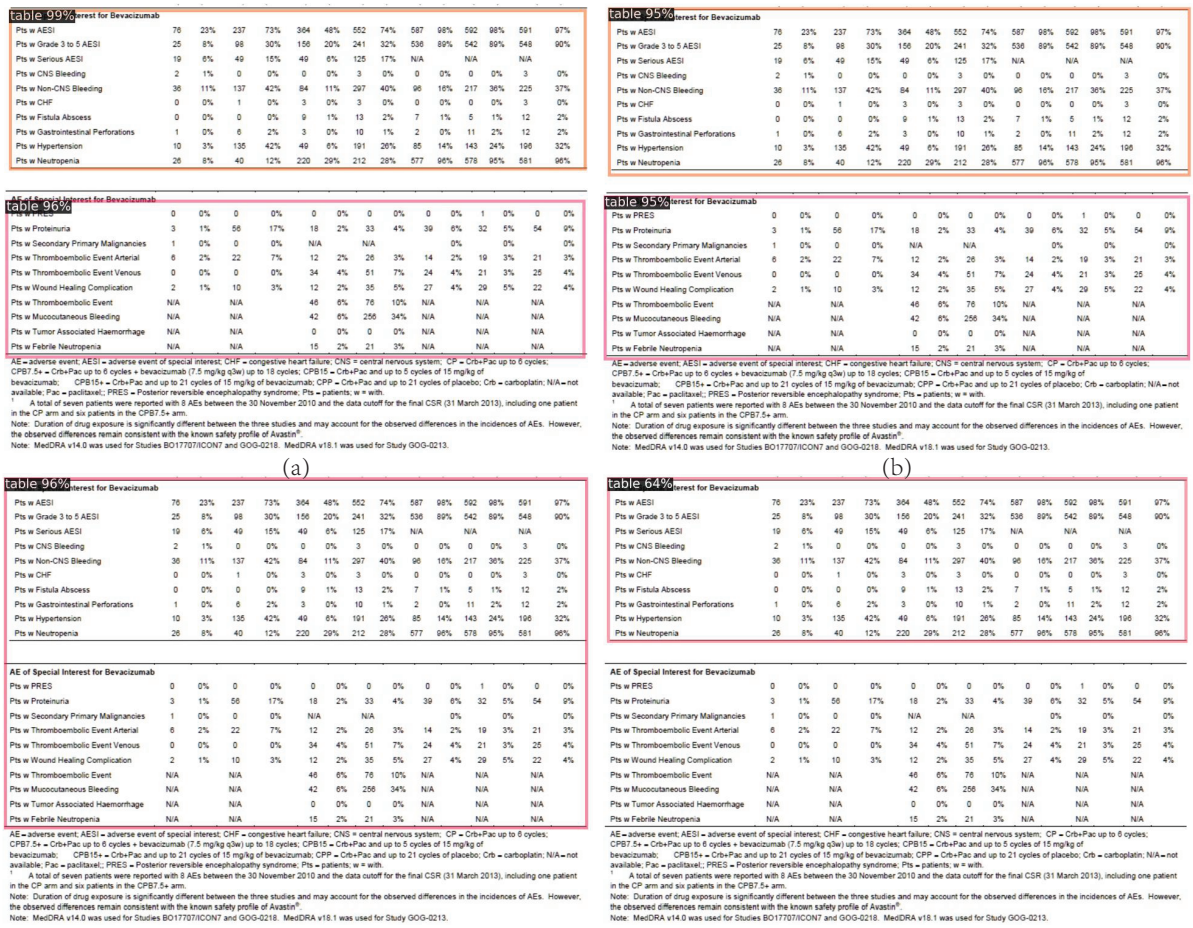



**图 7**：基线模型在 Open-Tables 测试集上的预测样本。图形 (a) (b) (c) (d) 分别是 TableDet、DiffusionDet、Deformable-DETR 和 SparseR-CNN 的结果。子图中的置信度分数分别为 99%、96%、95%、95%、96% 和 94%。

**图 8**：基线模型在 ICT-TD 测试集上的预测样本。图形 (a) (b) (c) (d) 分别是 TableDet、DiffusionDet、Deformable-DETR 和 SparseR-CNN 的结果。子图中的置信度分数分别为 100%、61%、99%、87%、82%、96%、96%、95% 和 87%。

与结果相比，性能大幅下降

在表格3 和表格 4。

## 4.3 噪声的影响

### Open-Tables 数据集

我们进行了额外实验，以讨论 Open-Tables 数据集中的噪声影响。在这些实验中，
Open-Tables 的训练集

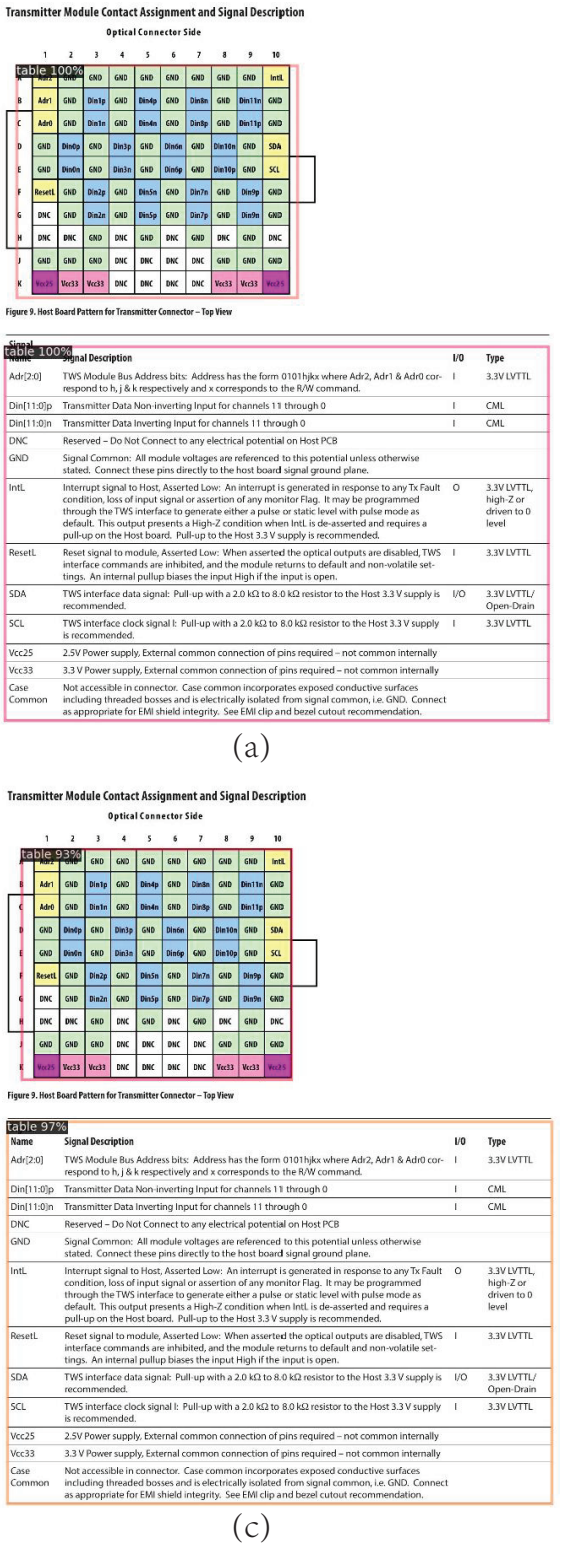



**图 9**：基线模型在 ICT-TD 测试集上的预测样本。图形 (a) (b) (c) (d) 分别为 TableDet、DiffusionDet、Deformable-DETR 和 SparseR-CNN 的结果。子图中的置信度分数分别为 100%、100%、96%、94%、93%、97%、73% 和 96%。

**表 4**：在 Open-Tables 数据集上的 F1分数实验结果。8834 和 1240 是样本数量。* 表示模型使用噪声样本进行训练。

| 数据集 | | 模型 | IoU 阈值下的 F1 分数 | | | | 加权平均F1 |
|---|---|---|---|---|---|---|---|
| 训练 | 测试 | | 80% | 85% | 90% | 95% | |
| Open-Tables 训练集 (8834) | Open-Tables 测试集 (1240) | TableDet | 96.8 | 95.1 | 92.6 | 83.9 | 91.8 |
| | | DiffusionDet | 97.8 | 96.7 | 93.8 | 84.5 | 92.9 |
| | | Deformable-DETR | 96.7 | 95.3 | 93.7 | 87.6 | 93.1 |
| | | SparseR-CNN | 97.5 | 95.9 | 93.4 | 87.4 | 93.3 |
| 噪声型 Open-Tables 训练集 (8834) | Open-Tables 测试集 (1240) | TableDet* | 96.1 | 94.7 | 92.3 | 81.6 | 90.8 |
| | | DiffusionDet* | 97.4 | 95.8 | 93.4 | 84.4 | 92.4 |
| | | Deformable-DETR* | 96.3 | 94.9 | 91.8 | 84.9 | 91.7 |
| | | SparseR-CNN* | 96.9 | 95.6 | 93.2 | 85.7 | 92.6 |

**表5**：跨域设置下的 F1分数实验结果。4000 和 1240 为样本数量。

| 数据集 | | 模型 | 不同交并比阈值下的 F1 分数 | | | | 加权平均F1 |
|---|---|---|---|---|---|---|---|
| 训练 | 测试 | | 80% | 85% | 90% | 95% | |
| ICT-TD 训练集 (4000) | Open-Tables 测试集 (1240) | TableDet | 77.8 | 74.2 | 68.6 | 52.1 | 67.6 |
| | | DiffusionDet | 85.3 | 80.8 | 72.6 | 53.7 | 72.4 |
| | | Deformable-DETR | 85.2 | 81.4 | 75.9 | 62.6 | 75.8 |
| | | SparseR-CNN | 80.0 | 76.3 | 69.7 | 55.8 | 69.9 |

使用含噪数据来训练基准模型，并使用 ICT-TD 数据集的测试集评估模型性能。如表 6 所示，清理后的 Open-Tables 数据集能够提升所有模型的性能，尤其是在交并比阈值高于 80% 时。实验结果验证了在创建 Open-Tables 数据集时进行噪声清理和标签对齐的必要性。此外，我们还使用清理后的 Open-Tables 测试集评估了这些模型，实验结果见表 4。与图 6 所示结果类似，使用清理后的 Open-Tables 训练集训练的模型表现更好比使用噪声型 Open-Tables 训练集训练的模型表现更好。值得一提的是，Open-Tables 测试集由 ICDAR2013、ICDAR2017、ICDAR2019 和 TNCR 数据集的清理后测试集合并而成，如表 1 所示。因此，表 4 中的结果也反映出，这些现有数据集能够从我们提出的 Open-Tables 数据集中获益。

## 4.4 潜在应用

在本节中，我们将讨论所提出的两个数据集的潜在应用。如前所述，ICT-TD 数据集是使用真实

**表6**：跨域设置下的 F1分数实验结果。8834 和 1000 表示训练样本数量。* 表示模型使用噪声样本进行训练。

| 数据集 训练 | 数据集 测试 | 模型 | 不同 IoU阈值下的 F1 分数 80% | 85% | 90% | 95% | 加权平均F1 |
|---|---|---|---|---|---|---|---|
| Open-Tables 训练集 (8834) | ICT-TD 测试集 (1000) | TableDet | 83.1 | 80.1 | 76.1 | 65.0 | 75.7 |
| | | DiffusionDet | 87.9 | 86.1 | 81.6 | 67.0 | 80.2 |
| | | Deformable-DETR | 84.1 | 82.2 | 79.7 | 70.1 | 78.7 |
| | | SparseR-CNN | 84.2 | 81.9 | 78.5 | 67.8 | 77.7 |
| 噪声型 Open-Tables 训练集 (8834) | ICT-TD 测试集 (1000) | TableDet* | 81.0 | 78.2 | 73.8 | 61.1 | 73.1 |
| | | DiffusionDet* | 86.2 | 84.0 | 79.8 | 66.3 | 78.6 |
| | | Deformable-DETR* | 81.9 | 80.4 | 77.9 | 67.7 | 76.7 |
| | | SparseR-CNN* | 85.0 | 82.5 | 78.4 | 64.2 | 77.1 |

ICT 领域的文档创建的。如表 6 所示，使用 Open-Tables 训练集训练模型，再在 ICT-TD 测试集上进行测试的跨域设置效果不佳。相比之下，如表3 7 所示，使用 ICT-TD 训练集训练的模型在 ICT-TD 测试集上测试时能够取得好得多的结果。因此，我们认为，所提出的 ICT-TD 数据集可用于训练和评估 ICT 领域专用模型，并可应用于 ICT 供应链优化问题 [46]，作为信息处理步骤的一部分。此外，所提出的 ICT-TD 数据集还可以丰富公开数据集的数据来源，并用于评估模型在跨领域设置中的泛化能力。另一方面，Open-Tables 数据集着重解决现有公开数据集中的噪声问题。如表 6 所示，清理后的 Open-Tables 训练集提高了模型在跨域设置中的泛化能力

设置。此外，Open-Tables 的测试集由清理后的 ICDAR2013、ICDAR2017、ICDAR2019 和 TNCR 数据集测试集合并而成，因此能够提供更可靠的评估结果。

## 5 结论

本文重新审视了一些具有高质量标注但标注定义不同的热门数据集，清理噪声样本，并统一这些数据集的标注，从而构建了一个更大且高质量的数据集，称为 Open-Tables。由于热门数据集的数据来源非常有限，我们使用 ICT 领域的数据表提出了一个新数据集，称为 ICT-TD。我们提出的 ICT-TD 数据集包含许多其他开放数据集中很少出现的领域特定样本，因此适用于跨领域设置。重新审视后的 Open-Tables 数据集具有一致性且规模更大，因而能够更可靠地评估模型性能。这

两个数据集可以作为更可靠的基准，用于构建可靠的表格检测应用；这类应用应避免丢失表格中的任何信息，并减轻噪声样本对模型评估的不利影响。最后，我们使用最先进的目标检测模型，为 ICT-TD 数据集和跨领域设置构建了强大的基线。实验结果表明，跨领域设置对表格检测问题而言更具挑战性。

现有大多数表格检测问题的研究都采用需要设定交并比阈值的目标检测评估指标。然而，这些评估指标无法直接反映从表格中提取信息的实际性能。例如，较大的预测框能够覆盖目标表格中的所有信息，即使其交并比分数较低，也优于交并比分数较高但可能遗漏目标表格部分信息的预测框。因此，采用其他指标评估模型，可以作为后续研究的一个方向，以弥补基于交并比分数的指标所存在的缺陷。

## 参考文献

[1] Y. Akkaya, M. Simsek, B. Kantarci, S. Khan，关于表格结构检测中裁剪与未裁剪训练集的研究。Neurocomputing**513**, 114–126 (2022)

[2] M. G¨obel、T. Hassan、E. Oro、G. Orsi、ICDAR 2013 年表格竞赛，载于*2013* 年第 *12* 届国际文档分析与识别会议(IEEE，2013)，第 1449–1453 页

[3] L. Gao, X. Yi, Z. Jiang, L. Hao, Z. Tang, ICDAR2017 页面目标检测竞赛，载于*2017 14thIAPR InternationalConference onDocument Analysisand Recog-nition (ICDAR)*，第1卷（IEEE，2017），第1417–1422页

[4]L. Gao、Y. Huang、H. D´ejean、J.L. Meu- nier、Q. Yan、Y. Fang、F. Kleber、E. Lang，Icdar 2019 表格检测与识别竞赛（CTDAR），载于*2019* 年国际文档分析与识别会议（*ICDAR*）（IEEE，2019），第 1510–1515 页

[5] M. Li、L. Cui、S. Huang、F. Wei、M. Zhou、Z. Li，Tablebank：用于基于图像的表格检测与识别的表格基准，载于第*12*届语言资源与评估会议论文集（2020），第1918–1925页

[6] X. Zhong、J. Tang、A.J. Yepes，《PubLayNet：迄今最大的文档版面分析数据集》，载于*2019* 年国际文档分析与识别会议（*ICDAR*）（IEEE，2019），第 1015–1022 页

[7]J. Fernandes、M. Simsek、B. Kantarci、S. Khan，TableDet：一种用于数据表图像中表格检测和表格图像分类的端到端深度学习方法。Neurocomputing **468**，317–334（2022）

[8]S. Chen, P. Sun, Y. Song, P. Luo, Diffusiondet：用于目标检测的扩散模型。arXiv预印本 arXiv:2211.09788（2022）[9] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai, Deformable detr：用于端到端目标检测的可变形变换器。arXiv预印本 arXiv:2010.04159（2020）[10] P. Sun, R. Zhang, Y. Jiang, T. Kong, C. Xu, W. Zhan, M. Tomizuka, L. Li, Z. Yuan, C. Wang, 等，Sparse r-cnn：基于可学习候选区域的端到端目标检测，载于*IEEE/CVF*计算机视觉与模式识别会议论文集（2021年），第14,454–14,463页[11]A. Abdallah, A. Berendeyev, I. Nuradin, D. Nurseitov, Tncr：表格网络检测与分类数据集。Neurocomputing **473**，79–97（2022）[12] A. Mondal, P. Lipps, C. Jawahar, Iiit-ar-13k：用于文档中图形对象检测的新数据集，载于文档分析系统：第*14*届*IAPR*国际研讨会，*DAS 2020*，中国武汉，*2020*年*7*月*26*日至*29*日，论文集 *14*（Springer，2020），第216–230页[13]J. Redmon, S. Divvala, R. Girshick, A. Farhadi, 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（2016），第779–788页[14] S. Ren, K. He, R. Girshick, J. Sun, Faster r- cnn：迈向实时目标检测，

区域候选网络。神经信息处理系统进展**28**（2015）[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg, Ssd：单次多框检测器，载于欧洲计算机视觉会议（Springer，2016），第 21–37 页[16]Z. Li, F. Zhou, Fssd：特征融合单次多框检测器。arXiv 预印本 arXiv:1712.00960（2017）[17] C. Ning, H. Zhou, Y. Song, J. Tang，用于目标检测的 Inception 单次多框检测器，载于*2017 IEEE* 多媒体与博览会研讨会国际会议（*ICMEW*）（IEEE，2017），第 549–554 页[18] M.J. Shafiee, B. Chywl, F. Li, A. Wong，Fast yolo：一种用于视频实时嵌入式目标检测的快速 You Only Look Once 系统。arXiv 预印本 arXiv:1709.05943（2017）[19]J. Redmon, A. Farhadi，Yolov3：渐进式改进。arXiv 预印本 arXiv:1804.02767（2018）[20]A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, Yolov4：目标检测的最佳速度与准确率。arXiv 预印本 arXiv:2004.10934（2020）[21] K. He, G. Gkioxari, P. Doll´ar, R. Girshick，Mask r-cnn，载于*IEEE* 国际计算机视觉会议论文集（2017），第 2961–2969 页

[22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszko- reit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polo- sukhin, Attention Is All You Need。神经信息处理系统进展**30**（2017年）

[23]X. Chu, Z. Tian, B. Zhang, X. Wang, X. Wei, H. Xia, C. Shen，视觉变换器的条件位置编码。arXiv 预印本 arXiv:2102.10882（2021年）[24] X. Dai, Y. Chen,J.Yang, P. Zhang, L. Yuan, L. Zhang，动态 DETR：采用动态注意力进行端到端目标检测，载于*IEEE/CVF* 国际计算机视觉会议论文集 （2021年），第 2988–2997 页[25] P. Gao, M. Zheng, X. Wang, J. Dai, H. Li，利用空间调制协同注意力实现 DETR 的快速收敛，载于*IEEE/CVF* 国际计算机视觉会议论文集 （2021年），第 3621–3630 页[26] Y. Huang, Q. Yan, Y. Li, Y. Chen, X. Wang, L. Gao, Z. Tang，基于 YOLO 的表格检测方法，载于*2019* 年文档分析与识别国际会议（*ICDAR*）（IEEE，2019年），第 813–818 页[27]D. Prasad, A. Gadpal, K. Kapadni, M. Visave, K. Sultanpure，Cascadetabnet：一种针对基于图像的文档的端到端表格检测与结构识别方法，载于*IEEE/CVF* 计算机视觉与模式识别会议论文集

识别研讨会 (2020)，第 572–573 页[28] Z. Cai、N. Vasconcelos，Cascade R-CNN：高质量目标检测与实例分割。IEEE 模式分析与机器智能汇刊**43**(5)，1483–1498（2019）[29]J. Wang、K. Sun、T. Cheng、B. Jiang、C. Deng、Y. Zhao、D. Liu、Y. Mu、M. Tan、X. Wang 等，面向视觉识别的深度高分辨率表征学习。IEEE 模式分析与机器智能汇刊**43**(10)，3349–3364（2020）[30]N. Carion、F. Massa、G. Synnaeve、N. Usunier、A. Kirillov、S. Zagoruyko，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（Springer，2020），第 213–229 页[31]B. Smock、R. Pesala、R. Abraham，PubTables-1M：迈向从非结构化文档中全面提取表格，载于*IEEE/CVF*计算机视觉与模式识别会议论文集（2022），第 4634–4642 页[32] S.A. Siddiqui、M.I. Malik、S. Agne、A. Dengel、S. Ahmed，Decnt：用于表格检测的深度可变形卷积神经网络。IEEE Access **6**，74,151–74,161（2018）[33]S. Schreiber、S. Agne、I. Wolf、A. Dengel、S. Ahmed，DeepDeSRT：用于表格检测与结构识别的深度学习

文档图像，载于*2017*年*IAPR*第*14*届文档分析与识别国际会议（*ICDAR*），第1卷（IEEE，2017），第1162–1167页[34]S.S. Paliwal、D. Vishwanath、R. Rahul、M. Sharma、L. Vig，《Tablenet：用于从扫描文档图像端到端进行表格检测和表格数据提取的深度学习模型》，载于*2019*年文档分析与识别国际会议（*ICDAR*）（IEEE，2019），第128–133页[35]E. Kara、M. Traquair、M. Simsek、B. Kantarci、S. Khan，《基于深度学习的数据表图像中表格结构发现的整体设计》。《人工智能工程应用》**90**，103,551（2020）[36] J. Jiang、M. Simsek、B. Kantarci、S. Khan，《Tabcellnet：基于深度学习的表格单元格结构检测》。Neurocomputing **440**，12–23（2021年）[37] J. Fang、X. Tao、Z. Tang、R. Qiu、Y. Liu，《用于表格检测评估的数据集、真实标注和性能指标》，载于*2012*年*IAPR*第*10*届文档分析系统国际研讨会（IEEE，2012），第445–449页[38] Z. Cai、N. Vasconcelos，《Cascade R-CNN：深入探究高质量目标检测》，载于*IEEE*计算机视觉与模式识别会议论文集(2018)，第 6154–6162 页[39]J. Song、C. Meng、S. Ermon，去噪扩散隐式模型。arXiv 预印本 arXiv:2010.02502（2020）[40] J. Ho、A. Jain、P. Abbeel，去噪扩散概率模型。神经信息处理系统进展**33**，6840–6851（2020）[41]Y. Wu、A. Kirillov、F. Massa、W.Y. Lo、R. Girshick。Detectron2。
h
ttps://github.co
m/facebookresearch/detectron2 （2019）[42] detrex 贡献者。detrex：一个用于基于 Transformer 的目标检测算法的研究平台。https://github.com/IDEA-Research/detrex （2022）[43]K. He、X. Zhang、S. Ren、J. Sun，面向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（2016），第 770–778 页[44] J. Deng、W. Dong、R. Socher、L.J. Li、K. Li、L. Fei-Fei，ImageNet：一个大规模分层图像数据库，载于*2009*年*IEEE*计算机视觉与模式识别会议（IEEE，2009），第 248–255 页[45]T.Y. Lin、M. Maire、S. Belongie、J. Hays、P. Perona、D. Ramanan、P. Doll´ar、C.L. Zit- nick，Microsoft COCO：通用物体text,in 计算机视觉——*ECCV 2014*：第*13*届

欧洲会议，瑞士苏黎世，*2014*年*9*月*6*日至*12*日，会议论文集，第*V13*部分(Springer, 2014)，第740–755页

[46]T. Lu、X. Guo、B. Xu、L. Zhao、Y. Peng、H. Yang，大数据领域的下一个重大议题：ICT 供应链的安全性，载于*2013* 年国际社会计算会议 (2013)，第 1066–1073 页。 https://doi.org/10.1109/SocialCom.2013.172

# 6 附录

## 6.1 详细实验结果

在本节中，我们列出详细的实验结果。具体而言，表格 7 和表格 8分别展示了 ICT-TD 和 Open-Tables 数据集上的结果。表格9和表格 10 展示了两种跨域设置的结果：第一种设置使用 ICT-TD 数据集的训练集和 Open-Tables 数据集的测试集；第二种设置使用 Open-Tables 数据集的训练集和 ICT-TD 数据集的测试集。

**表 7**：ICT-TD 数据集的详细实验结果。

| 方法 | 指标 | IoU | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50% | 55% | 60% | 65% | 70% | 75% | 80% | 85% | 90% | 95% | 50%:95% |
| TableDet | 精确率 | 97.0 | 96.1 | 96.1 | 95.9 | 94.9 | 94.0 | 93.0 | 91.6 | 88.5 | 73.2 | 92.0 |
| | 召回率 | 98.0 | 97.6 | 97.5 | 97.2 | 96.5 | 95.8 | 94.8 | 93.3 | 90.6 | 78.9 | 94.0 |
| | F1 | 97.5 | 96.8 | 96.8 | 96.5 | 95.7 | 94.9 | 93.9 | 92.4 | 89.5 | 75.9 | 93.0 |
| DiffusionDet | 精确率 | 96.9 | 96.7 | 96.5 | 96.2 | 95.2 | 94.7 | 93.8 | 92.4 | 89.4 | 73.8 | 92.6 |
| | 召回率 | 99.2 | 99.0 | 98.9 | 98.8 | 98.4 | 97.9 | 97.2 | 96.1 | 93.2 | 79.4 | 95.8 |
| | F1 | 98.0 | 97.8 | 97.7 | 97.5 | 96.8 | 96.3 | 95.5 | 94.2 | 91.3 | 76.5 | 94.2 |
| 可变形-DETR | 精确率 | 97.0 | 96.7 | 96.4 | 96.0 | 95.1 | 94.4 | 93.8 | 92.4 | 90.0 | 79.4 | 93.1 |
| | 召回率 | 98.9 | 98.8 | 98.6 | 98.4 | 97.4 | 96.9 | 96.5 | 95.2 | 93.2 | 84.9 | 95.9 |
| | F1 | 97.9 | 97.7 | 97.5 | 97.2 | 96.2 | 95.6 | 95.1 | 93.8 | 91.6 | 82.1 | 94.5 |
| SparseR-CNN | 精确率 | 95.8 | 95.5 | 95.2 | 95.0 | 94.3 | 93.6 | 92.6 | 91.1 | 88.4 | 75.6 | 91.7 |
| | 召回率 | 98.7 | 98.4 | 98.2 | 98.1 | 97.5 | 97.2 | 96.1 | 94.7 | 92.5 | 83.4 | 95.5 |
| | F1 | 97.2 | 96.9 | 96.7 | 96.5 | 95.7 | 95.4 | 94.3 | 92.9 | 90.4 | 79.3 | 93.6 |

**表格 8**：Open-Tables 数据集上的详细实验结果。

| 方法 | 指标 | IoU | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50% | 55% | 60% | 65% | 70% | 75% | 80% | 85% | 90% | 95% | 50%:95% |
| TableDet | 精确率 | 98.7 | 97.9 | 97.9 | 97.8 | 97.8 | 96.7 | 96.6 | 94.3 | 91.7 | 81.8 | 95.1 |
| | 召回率 | 99.1 | 99.0 | 98.9 | 98.5 | 98.4 | 97.7 | 97.1 | 95.9 | 93.5 | 86.1 | 96.4 |
| | F1 | 98.9 | 98.4 | 98.4 | 98.1 | 98.1 | 97.2 | 96.8 | 95.1 | 92.6 | 83.9 | 95.7 |
| DiffusionDet | 精确率 | 98.3 | 98.2 | 98.0 | 97.9 | 97.8 | 97.2 | 96.8 | 95.3 | 92.0 | 81.8 | 95.3 |
| | 召回率 | 99.8 | 99.7 | 99.7 | 99.7 | 99.6 | 99.4 | 98.9 | 98.1 | 95.6 | 87.4 | 97.8 |
| | F1 | 99.1 | 99.0 | 98.8 | 98.8 | 98.7 | 98.3 | 97.8 | 96.7 | 93.8 | 84.5 | 96.6 |
| 可变形-DETR | 精确率 | 97.8 | 97.7 | 97.7 | 97.5 | 97.0 | 96.5 | 95.8 | 94.1 | 92.0 | 85.2 | 95.1 |
| | 召回率 | 99.4 | 99.3 | 99.2 | 99.0 | 98.6 | 98.0 | 97.6 | 96.6 | 95.3 | 90.3 | 97.3 |
| | F1 | 98.6 | 98.5 | 98.4 | 98.3 | 97.8 | 97.2 | 96.7 | 95.3 | 93.7 | 87.6 | 96.2 |
| SparseR-CNN | 精确率 | 98.4 | 98.2 | 98.1 | 97.9 | 97.7 | 97.0 | 96.3 | 94.2 | 91.4 | 84.5 | 95.4 |
| | 召回率 | 99.7 | 99.7 | 99.7 | 99.7 | 99.5 | 99.1 | 98.7 | 97.7 | 95.5 | 90.5 | 98.0 |
| | F1 | 99.1 | 99.0 | 98.9 | 98.8 | 98.6 | 98.0 | 97.5 | 95.9 | 93.4 | 87.4 | 96.7 |

**表9**：跨域设置下的详细实验结果。训练集来自 ICT-TD，测试集来自 Open-Tables。

| 方法 | 指标 | IoU | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50% | 55% | 60% | 65% | 70% | 75% | 80% | 85% | 90% | 95% | 50%:95% |
| TableDet | 精确率 | 88.4 | 87.2 | 86.0 | 84.9 | 82.8 | 79.4 | 75.7 | 72.0 | 66.4 | 48.6 | 77.1 |
| | 召回率 | 91.9 | 90.4 | 89.4 | 88.3 | 86.6 | 83.7 | 80.1 | 76.6 | 71.0 | 56.0 | 81.4 |
| | F1 | 90.1 | 88.8 | 87.7 | 86.6 | 84.7 | 81.5 | 77.8 | 74.2 | 68.6 | 52.1 | 79.2 |
| DiffusionDet | 精确率 | 86.2 | 85.5 | 84.8 | 83.7 | 82.3 | 80.5 | 77.5 | 72.7 | 65.3 | 48.5 | 76.7 |
| | 召回率 | 98.8 | 98.4 | 98.3 | 97.9 | 97.4 | 96.6 | 94.9 | 91.0 | 81.8 | 60.1 | 91.5 |
| | F1 | 92.1 | 91.5 | 91.1 | 90.2 | 89.2 | 87.8 | 85.3 | 80.8 | 72.6 | 53.7 | 83.4 |
| 可变形-DETR | 精确率 | 91.2 | 90.4 | 89.5 | 88.3 | 86.6 | 84.1 | 80.9 | 76.9 | 71.3 | 57.3 | 81.6 |
| | 召回率 | 97.3 | 96.7 | 96.0 | 95.4 | 94.1 | 93.0 | 90.0 | 86.5 | 81.2 | 69.0 | 89.9 |
| | F1 | 94.2 | 93.4 | 92.6 | 91.7 | 90.2 | 88.3 | 85.2 | 81.4 | 75.9 | 62.6 | 85.5 |
| SparseR-CNN | 精确率 | 85.6 | 84.6 | 83.4 | 81.9 | 80.3 | 78.0 | 74.5 | 70.8 | 64.7 | 50.9 | 75.5 |
| | 召回率 | 97.2 | 96.3 | 95.2 | 93.9 | 92.1 | 89.9 | 86.4 | 82.7 | 75.5 | 61.8 | 87.1 |
| | F1 | 91.0 | 90.1 | 88.9 | 87.5 | 85.8 | 83.5 | 80.0 | 76.3 | 69.7 | 55.8 | 80.9 |

**表 10**：跨领域设置下的详细实验结果。训练集来自 Open-Tables，测试集来自 ICT-TD。

| 方法 | 指标 | IoU | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 50% | 55% | 60% | 65% | 70% | 75% | 80% | 85% | 90% | 95% | 50%:95% |
| TableDet | 精确率 | 86.5 | 86.2 | 85.2 | 85.0 | 84.0 | 83.6 | 81.8 | 78.7 | 74.4 | 62.3 | 80.8 |
| | 召回率 | 88.9 | 88.4 | 87.7 | 87.3 | 86.7 | 86.1 | 84.5 | 81.5 | 77.9 | 67.9 | 83.7 |
| | F1 | 87.7 | 87.3 | 86.4 | 86.1 | 85.3 | 84.8 | 83.1 | 80.1 | 76.1 | 65.0 | 82.2 |
| DiffusionDet | 精确率 | 88.1 | 87.4 | 86.9 | 86.5 | 86.0 | 85.1 | 83.7 | 82.0 | 77.6 | 63.8 | 82.7 |
| | 召回率 | 95.8 | 95.3 | 95.2 | 94.8 | 94.3 | 93.9 | 92.5 | 90.7 | 86.2 | 70.6 | 90.9 |
| | F1 | 91.8 | 91.2 | 90.9 | 90.5 | 90.0 | 89.3 | 87.9 | 86.1 | 81.6 | 67.0 | 86.6 |
| 可变形-DETR | 精确率 | 88.4 | 87.7 | 87.0 | 85.7 | 83.5 | 81.9 | 80.2 | 78.7 | 76.3 | 66.3 | 81.6 |
| | 召回率 | 94.5 | 94.0 | 93.4 | 92.8 | 91.1 | 89.9 | 88.4 | 86.2 | 83.3 | 74.4 | 88.8 |
| | F1 | 91.3 | 90.7 | 90.1 | 89.1 | 87.1 | 85.7 | 84.1 | 82.2 | 79.7 | 70.1 | 85.0 |
| SparseR-CNN | 精确率 | 84.5 | 83.9 | 83.4 | 82.9 | 81.8 | 80.8 | 79.7 | 77.4 | 73.9 | 63.4 | 79.2 |
| | 召回率 | 94.9 | 94.4 | 94.1 | 93.5 | 92.5 | 91.2 | 89.4 | 86.9 | 83.6 | 72.8 | 89.3 |
| | F1 | 89.4 | 88.8 | 88.4 | 87.9 | 86.8 | 85.7 | 84.2 | 81.9 | 78.5 | 67.8 | 83.9 |