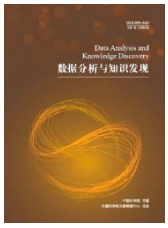


数据分析与知识发现
Data Analysis and Knowledge Discovery
ISSN 2096-3467,CN 10-1478/G2

《数据分析与知识发现》网络首发论文

题目：基于 YOLOv5-ECA-BiFPN 的学术期刊文献图表识别与提取方法研究

作者：李英群，李亚菲，裴雷，胡志伟，宋宁远

网络首发日期：2023-02-07

引用格式：李英群，李亚菲，裴雷，胡志伟，宋宁远．基于 YOLOv5-ECA-BiFPN 的学术期刊文献图表识别与提取方法研究[J/OL]．数据分析与知识发现.
https://kns.cnki.net/kcms/detail//10.1478.G2.20230206.1845.004.html

网络首发时间：2023-02-07 10:14:43
网络首发地址：https://kns.cnki.net/kcms/detail//10.1478.G2.20230206.1845.004.html

基于 YOLOv5-ECA-BiFPN 的学术期刊文献图表识别与提取方法研究*

李英群[1,2]，李亚菲[1,2]，裴雷[1,2]，胡志伟[1,2]，宋宁远[1,2]
[1](南京大学信息管理学院 南京 210023)
[2](南京大学数据智能与交叉创新实验室 南京 210023)

摘要：[目的]精准识别与提取学术期刊文献中的图表，促进学术图表的传播和交流。**[方法]**将 YOLOv5 算法中引入 ECA 通道注意力模块，并优化 PAN 模块为 BiFPN，随机抽样十三个学科门类 1300 篇学术期刊文献作为实验数据，利用 poppler-0.68.0 将其转换为高质量的图片，并基于该数据集验证新算法性能。**[结果]**相较于次优值，新模型数据集 F1 值提高 1.99%，达到 99.88%。**[局限]**数据标注范围与数量有待扩大，可覆盖至更多场景。**[结论]**基于 YOLOv5-ECA-BiFPN 的方法能够有效提高特殊场景下的图表识别与提取效果。
关键词：学术期刊文献；YOLOv5-ECA-BiFPN；学术图表
分类号：TP391.1，G256
DOI:

The identification and extraction of figures and tables in academic journal literature based on YOLOv5-ECA-BiFPN

Li Yingqun[1,2], Li Yafei[1,2], Pei Lei[1,2], Hu Zhiwei[1,2], Song Ningyuan[1,2]
[1](School of Information Management, Nanjing University, Nanjing 210023, China)
[2](Laboratory of Data Intelligence and Cross Innovation, Nanjing University, Nanjing 210023, China)

Abstract：[Objective] In order to promote the dissemination of academic figures and tables, it is of great significance to accurately identify and extract figures and tables from academic literature. **[Methods]**The ECA channel attention module is introduced into the YOLOv5 algorithm, and the PAN module is replaced with BiFPN. 1300 academic journal literature is then randomly selected

通讯作者(Corresponding author)：李亚菲(Li Yafei)，ORCID：0000-0003-1754-2300，Email:dg20140013@smail.nju.edu.cn。

*本文系 2022 年度南京大学数据智能与交叉创新实验室研究项目《去中心化智能数据引擎设计研究》阶段性成果。
The work is supported by “Research on the Design of Decentralized Intelligent Data Engines”, a research project of Laboratory of Data Intelligence and Cross Innovation of Nanjing University,2022.

among thirteen subject categories as experimental data and converted into high-quality images using poppler-0.68.0. The performance of the new algorithm is verified based on this dataset. **[Results]**Compared with the suboptimal algorithm, the F1 value of the new model is improved by 1.99% to 99.88% when applied to the dataset. **[Limitations]** The scope and quantity of data annotation needs to be expanded, so that more situation can be covered. **[Conclusions]** YOLOv5-ECA-BiFPN can effectively improve the effect of recognition of figures and tables in special situation.

Keywords：Academic journal literature; YOLOv5-ECA-BiFPN; Academic figures and tables

1 引言

学术期刊文献反映了各学科领域最新研究成果和发展动态，是学术交流的重要载体。而学术图表作为文献中信息集聚的典型视觉资源，一般具有内容概括、论点支撑、数据展示等功能[1]，成为文献阅读或浏览中优先关注的功能区间和视觉搜索区域。因此，学术图表的识别提取、语义加工和智能应用越来越受到科技情报届的关注。

国内外学术文献格式有 PDF、CAJ、HTML、XML、EPUP 等，其中 PDF 格式是较为流行的格式，并分为结构化与非结构化两种[2]。面向结构化文献，已有研究多采取标签化识别[3]、分析组件[4]、解析 PDF 文件[5]的方式进行图表的识别与提取，但噪音图像多，矢量图像嵌合类型复杂，标签信息与本体语义存在偏差的问题尚且无法解决。面向非结构化文献，已有研究常采用图像分割[6]、像素点连通[7]等方法对图表进行提取，但图表像素低、分布不均匀、图像构成多样化等因素造成基于此类方法提取时发生错误，导致召回率偏低或图表信息缺失。上述方式本质上均依据页面解析、元素拆解、规则梳理等方式对数据集进行分析，缺少基于图表本体语义信息的提取与融合，导致特殊场景下的图表识别能力较弱，只能拘泥于同质化程度高、排版结构相似的文献进行识别提取，难以支撑规模性、连贯性研究任务的数据需求，分散在不同学术期刊文献中的图表成为了无法进一步挖掘利用的“数据孤岛”。

本文以 PDF 格式学术期刊文献中的图表为研究对象，提出一种改进 YOLOv5 的学术图表识别与提取方法，通过深度学习技术提取图表特征信息与本体语义，以一种通用性强、泛化能力高的方式有效识别与提取不同类型 PDF 学术期刊文献中的图表，解决噪音图像多、矢量图像构成复杂、图表像素低等特殊场景下的图表提取问题。

2 相关研究

学术期刊文献中图表识别与提取的研究主要根据 PDF 文献类型展开。在结构化文献中，图像主要以 raster 栅格（JPEG，PNG）或 vector formats 矢量格式（EPS，SVG）嵌入于 PDF 文献中，表格主要是文本与矢量线性元素的混合结构。其中，raster 栅格图表（JPGE，PNG）通常以单独的内容主体（XObject）嵌入到 PDF 文献中。XPDF[5]、Apache PDFBox[8]和 PDFminer[9]等 PDF 解析工具可以有效地进行图表的提取，但学术期刊文献中嵌入的“噪音”图片会对提取结果产生干扰，如封面上的出版社 LOGO、页眉处的装饰性图案、文末的作者照片等。矢量图像在绘制过程中往往包含大量的矢量元素与嵌入的栅格图像[10]，而基于上述的 PDF 解析工具提取时只能提取矢量图像的单个组件，如散点

图中的分布点或线形图中的某一条线段，无法有效地识别图表整体。为此，部分研究者提出启发式算法[11]来识别图像标题与辅助信息，随后利用分类算法[3,12]、聚类算法[13]、深度神经网络[14]等方式连接组件或排除无效元素。Li[15]等人首先将文本和图形内容分开，然后利用布局信息来有效地检测和提取图像和标题，最终建立起 PDFigCapX 图像提取系统，自动生成包含图像及其相关标题的文件。但标题信息有时与图表本体语义存在一定的偏差或缺失，图像的构成可能出现不同元素间距过大，导致后续处理时可能出现分类算法将其视为两张图像的异常提取情况。

在非结构化文献中，页面内容主要是通过早期纸质学术期刊文献经扫描后获得。当前此类文献图表的识别与提取主要采用图像分割或像素点连通的方法。Chen K[16]与 Amin A[17]均采取页面图像分割的方式，分析内容区域和空白区域的差异，随后确定图表位置。Choudhury S R[6]首先将结构化的文献中的文本信息剔除，随后将页面转化为图像格式后利用图像分割进行识别。Gomez-Kramer[18]通过比较不同纹理特征的分割性能，提出一种多纹理特征的评估方案，可以更好地实现图像与文字的分离。但图像构成形式多样，无法明确所有图像区域的边界，甚至会出现图像切分过度的情况。Simon A[7]与 Ha J[19]则通过计算像素点之间的连通关系，结合一定的规则合并相邻内容，最终确定图表的位置。但时间久远的学术期刊文献主要以扫描件为主，图像绘制工具单一，存在大量的低像素图表，仅通过像素连通方式无法实现该类图表的提取。

上述方法均可以在一定程度上实现学术期刊文献图表的识别与提取，但方法的选择依赖于 PDF 学术期刊文献本身的格式，同时提取方式着重于图表的嵌入形式与周边特征信息的规则梳理，通常以剥离非可视化信息或连接图像组件的方式逐步确定图像区域，忽略了图表本身所包含的语义信息；而表格的提取则采用线框结构识别[20]与整体布局分析[21]，但方法的选取受限于文献类型[22]。此外，矢量图表的提取通常统一处理，难以做到图像与表格的精确化区分。目前我国学术期刊文献跨度时间长、内容排版多样、早期图表使用不规范，造成图表结构形式多样、矢量图像嵌合类型复杂、标签信息与本体语义存在偏差等问题；同时受限于版权保护的原因，同一 PDF 文献中会出现结构化与非结构化页面交叉布局的情况。已有研究通常局限于小部分同质化高的学术期刊文献，利用人工的方式针对图表分布的位置特征进行规则梳理，暂未提出通用性的识别与提取方法。本文构建以 YOLOv5 为基础的目标检测算法，通过深度学习的方式着眼于图表本体语义信息的抽取和融合，有效提高多场景下的图表识别与提取效果。

3 研究框架及算法

3.1 研究框架

本研究主要包括三个主要步骤，如图 1 所示。

第一，原始数据集转化与制备。通过随机抽样从 13 个一级学科的核心期刊中抽取 1300 篇 PDF 格式的文献，选用 poppler-0.68.0 将文献页面转换为 11590 张高质量图片作为原始数据集，同时进行图像命名和页数识别。

第二，图像标记与格式转化。基于深度学习图像标注平台 labelImage 对数据集进行语义分割和人工标注，并将标注后的 XML 格式数据转化为 YOLO 支持的 TXT 格式。

第三，算法实验。结合学术期刊文献图表的特征与分布特点，本文采用 YOLOv5 算法，主要对骨干网络架构（Backbone）和颈部（Neck）网络架构进行优化：首先在骨干网络（Backbone）中引入 ECA 通道注意力机制，以聚焦位置感知与方向感知信息，加强对图表核心特征的关注度；随后基于路径增强的思想，面向特征金字塔结构进行改进，以一种加权双向网络结构在 PAN 的基础上优化为 BiFPN，通过双向跨尺度连接实现底层目标位置信息与高层特征语义信息相融合，改变算法原有信息流传递方式，以减少计算过程中的信息损耗；最后通过消融实验对不同的优化情况效果进行精度测量。

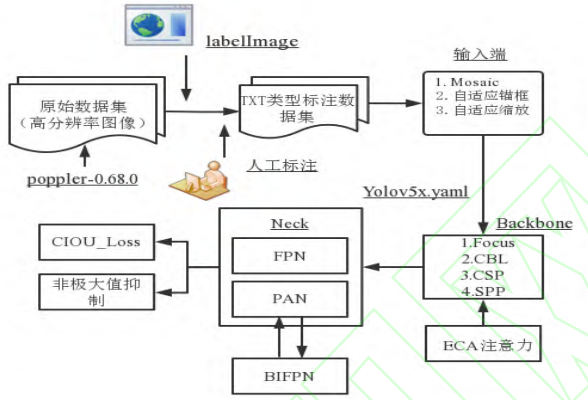


图 1　技术框架图

Fig.1 Technical Framework

3.2　YOLOv5 算法框架

YOLOv5 是一种基于计算机视觉技术与深度学习技术，确定目标图像中物体的类别与位置的算法。YOLOv5 引入 Focus 模块对特征图进行切片、连接，以获取更多特征信息；随后利用 CBL、C3 等模块进行特征提取，通过 SPP 整合后输入 Neck 网络端中，特征金字塔网络（FPN）和金字塔注意力网络（PAN）进行特征融合，实现图表浅层网络的特征信息和深层网络的语义信息有效结合，有效地避免背景错误，加强算法的预测能力。整体网络架构如图 2 所示，其网络结构分为输入端、Backbone、Neck、Prediction 四个部分。

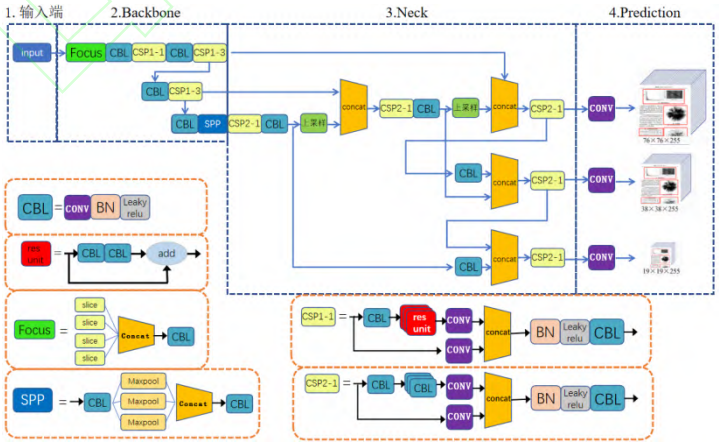


图 2　YOLOv5 算法网络结构[23]

Fig.2 YOLOv5 Algorithm Network Structure[23]

输入端主要包括 Mosaic 数据增强、自适应锚框计算、自适应图片缩放三个部分，以实现丰富图像背景、迭代网络参数、减少训练过程中的信息冗余等作用。Backbone 特征提取网络是在不同的图像细粒度上聚合形成图像特征的卷积神经网络，如图 3 所示，Backbone 网络首先通过 Focus 模块将 3 通道的图像切片为 12 通道，随后使用 concat 操作从深度上连接这 4 个切片，通过卷积层后得到相关采样特征图，减少了计算量的同时又提升速度，使得模型学习到更多的特征。Neck 中采用了特征金字塔网络（FPN）[24]与金字塔注意力网络（PAN）[25]相结合的结构，将常规的 FPN 层与自下而上的特征金字塔进行结合，同时融合提取的语义特征与位置特征，并将主干层与检测层特征融合，使模型能够提取更加丰富的特征信息。Prediction 输出的结果向量主要涉及目标对象的类别概率、对象得分和该对象边界框的位置等信息，检测网络由三层检测层组成，不同尺寸的特征图用于检测不同尺寸的目标对象。

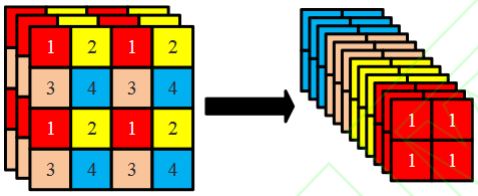


图 3　Focus 切片操作

Fig.3 Focus Slicing Operation

3.3　ECA 通道注意力模块

低像素图表本体信息划分不显著，会与图像背景环境分割不明显，而原始的 YOLOv5 算法在进行卷积采样时会丢失部分此类图表的特征信息。此外，图表类型愈发多样化，复合图像、多元化结构图可能导致算法在识别时出现切分过度，有些以文本元素构成的图像进一步加大了算法在识别时背景信息的干扰。为解决上述问题，本文在 YOLOv5 算法的 backbone 结构中引入 ECA 通道注意力模块[26]，将计算资源更多地聚集在图表本体信息，增强网络对图表核心特征的关注度，并抑制干扰性特征，有效避免背景信息的干扰，以解决低像素点、图表构成多样化的问题。

ECA 的工作机制如图 4 所示。首先针对输入进来的特征层进行全局平均池化；随后将池化后的特征长条进行 1D 卷积提取，通过 sigmoid 函数生成每一个特征点的权值；最后再与原始特征层结合，便可以获得具有通道注意力的特征层。本文通过修改 backbone，将原 C3 模块后加入该模块，以提升算法性能。

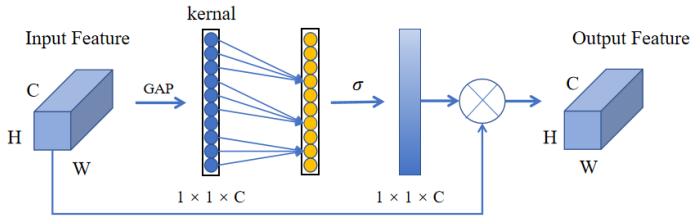


图 4　ECA 模块[26]

Fig.4 ECA Module[26]

3.4　BiFPN 模块

YOLOv5 算法中的 FPN 结构（图 5（a））建立了一条自上而下的通道融合

结构，以实现图像语义层面与特征层面的有效融合，但这种结构会受到单向信息流的限制。图 5（b）是为了解决 FPN 结构问题所提出的一种新的 PAN 结构，它在 FPN 结构的基础上添加了一条自下而上的通道，可以将底层的图像信息都传输至预测特征层中，从而使得预测层同时包含顶层语义和底层特征信息，但 PAN 增加了网络模型复杂度，降低了信息传递效率。

因此，本文在 PAN 的基础之上优化为 BiFPN[27]，如图 5（c）所示。BiFPN 结构是以 PAN 为基础，将无特征融合且贡献度较小的节点进行删除，并在原始输入节点和输出节点之间新增通道，从而在节省资源消耗的同时融合更多的特征信息。BiFPN 结构通过权值与权值之和之比来进行快速归一化融合，最终将权值归一到[0,1]之间，提高对不同情况下目标的感知能力，在预测端可以融合不同层级间特征图的信息，有效解决噪音图像等因素带来的干扰。

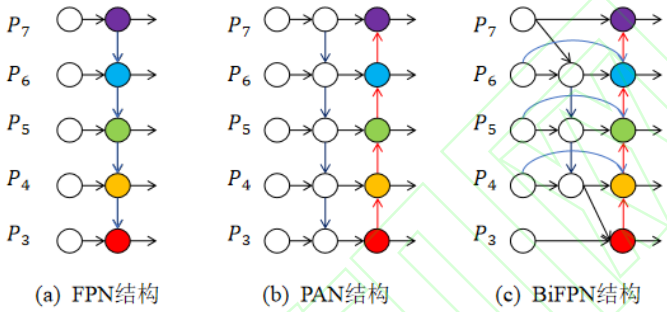


图 5　FPN、PAN 和 BiFPN 结构[27]

Fig.5 FPN、PAN and BiFPN Structure[27]

4　实验

4.1　数据集

（1）数据选择及预处理

为保证 YOLOv5-ECA-BiFPN 算法对学术期刊文献图表识别的准确率和泛化力，本文以 13 个一级学科门类为基础，从 CNKI 中随机选取并下载 1982-2022 年的 1300 篇学术期刊文献作为研究对象，以 poppler-0.68.0 将 PDF 文件转化为分辨率为 300dpi 的图片，文件每一页转化为一个存储单位图片，以对应页码为图片名称后缀，最终生成 11590 张 JPG 图片数据，详情如图 6 所示。

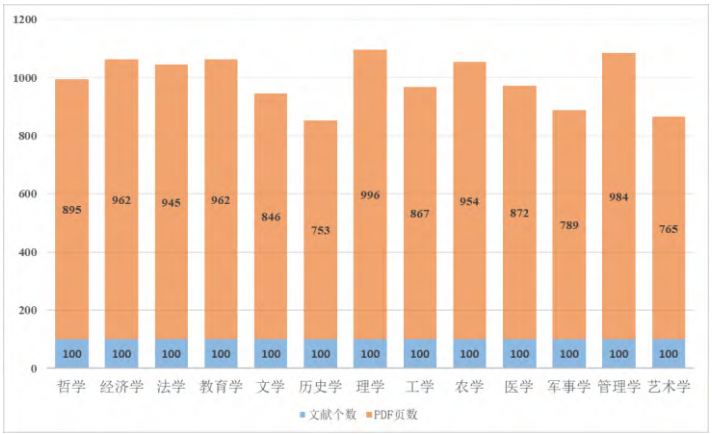


图 6　学术图表识别与提取数据集

Fig.6 Scientific Research Chart Recognition and Extraction Data Set

（2）数据集制备

本文将收集到的图片数据进行筛选和处理，使用深度学习图像标注软件 LabelImage 对样本数据中的图片进行标注，每张图片上的所有图像和表格使用最小的矩形框标定，分别对应 figure 与 table 标签集，数据标注样式如图 7 所示。

图 7　LabelImage 标注示例图

Fig.7 LabelImage Example Drawing of Annotation

本文将标注好的数据集保存为 PASCAL VOC 格式的 XML 文件，其中包含目标识别物体的类别与目标检测框坐标信息。本文通过如下公式将其转化为 YOLO 可识别的 TXT 格式的数据文献：

$$x = (x_{\max} + x_{\min}) / d_w \quad (1)$$

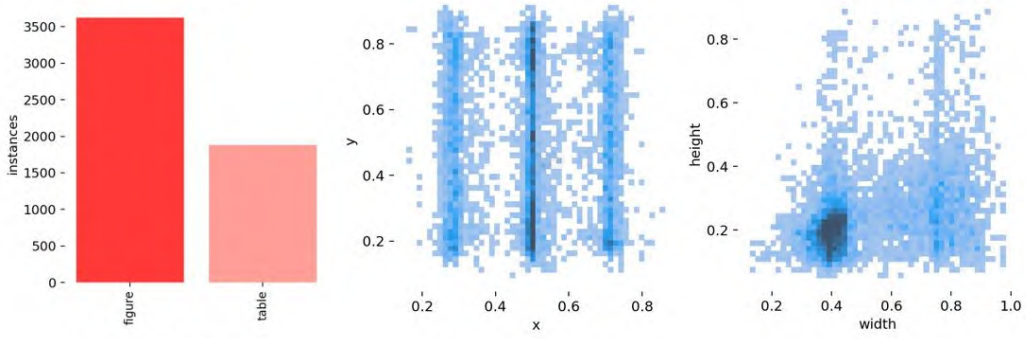
$$y = (y_{\max} + y_{\min}) / d_h \quad (2)$$

$$w = (x_{\max} - x_{\min}) / d_w \quad (3)$$

$$h = (y_{\max} - y_{\min}) / d_h \quad (4)$$

其中(x_{max},y_{max})，(x_{min},y_{min})分别表示目标检测框对应的左上角与右下角的坐标信息，分别表示检测框的宽度与高度信息。本文通过编写脚本进行批量地操作和处理，将 XML 格式文件转化为 TXT 文件，随后将数据集以 8:1:1 的比例分为训练集、测试集、验证集。如图 8 所示，本文统计了标注数据集上的图表数量，并针对数据集中目标框的大小与位置分布进行了数据分析与可视化研究，将目标框中心点的坐标与宽度高度做归一化处理，刻画目标框位置与大小分布图。因标签集在整体上有着较好的代表性，所以图 8 在刻画数据集的同时，在一定程度上可以揭示目前学术期刊文献中图表的使用情况。如图 8（a）显示，图像标签的数量远高于表格的数量，在学术期刊文献中图表的使用更加频繁，这与图表对抽象概念的具现化密不可分；如图 8（b）所示，目标框中心分布图呈现三条状分布的特点，其中较多的图表中心位于页面中下部与中上部；如图 8（c）显示，图表的宽度往往大于高度，宽高比例主要位于 1.265-

1.758 区间范围内。



(a) 标签数量统计图　(b) 图表中心分布图　(c) 图表大小分布图

图 8　标签数据集信息分布图

Fig.8 Label Dataset Information Distribution

4.2　实验设计

（1）部署环境与参数设置

深度学习实验需要一定高性能的计算算力与诸多开源软件和框架的支持，本实验中数据集制作、模型训练工作均于同一计算机上进行，具体计算资源配置如表 1 所示。

表 1　实验软硬件环境配置

Table1 Experimental Hardware Environment

类别	参数
系统	Windows10
CPU	Intel Core i7 9700K
GPU	NVIDIA GeForce RTX2080Ti
固态硬盘	500GB
Python	Python 3.8
CUDA	CUDA 11.3
PyTorch	PyTorch 1.11
OpenCV	OpenCV 4.3.2

本实验以预训练模型 yolov5x.pt 进行权重初始化，设置图片标准尺寸为 640×640×3，动量因子为 0.937，色调、饱和度、明度的增强系数分别为 0.015、0.7、0.4，批处理尺寸大小为 8，训练 epochs 为 80。模型训练结束后保存相关权重文件，同时针对测试集上针对模型的基本情况进行评估。

（2）算法评价指标

本文选取F_1、平均精度均值(mAP)、损失函数来针对模型训练效果进行评价。其中F_1值是精确率（Precision）和召回率（Recall）加权调和平均，是结合两者提出的综合性评估指标，可以较好地避免两者相差较大而产生的问题。精确率又称为查准率，是指算法预测出所有的正样本中，实际正确的图表占比；召回率又称为查全率，是算法检测正确结果占比图表中所有目标的比例。具体计算公式如下：

$$F_1 = 2\frac{Prcision * Recall}{Precision + Recall}$$

$$Recall = \frac{TP}{TP + FP}$$

$$Precision = \frac{TP}{TP + FP}$$

AP（Average Precision）能够反映单个目标类别的检测性能，mAP（mean Average Precision）定义为所有类别 AP 的均值，越高意味着模型越容易在高召回率下保持高准确率，计算公式如（8）、（9）所示。

$$AP = \int_0^1 P(R)dR$$

$$mAP = \frac{1}{C}\sum_{c \in C} AP(C)$$

YOLOv5 算法的损失函数可以分为三个部分：置信度损失、分类损失和目标框定位损失，整体计算公式如（10）所示。

$$LOSS = l_{obj} + l_{cls} + l_{box}$$

本文所采用的是 CIOU，CIOU 损失函数在 DIOU 的基础之上考虑了边界框中心点距离的信息与边界框宽高比的尺度信息，同时也考虑了两框之间的长宽比，这使得检测框定位的更加准确，提高了模型的检测性能。CIOU 损失函数计算公式如（11）、（12）、（13）所示。

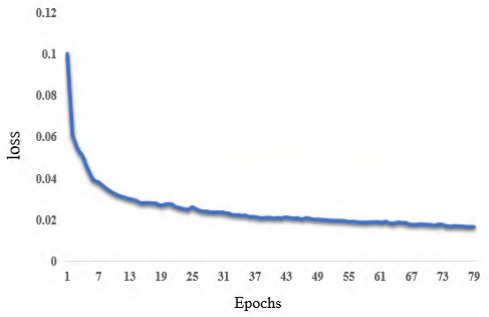
$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha\upsilon$$

$$\alpha = \frac{\upsilon}{(1 - IoU)} + \upsilon$$

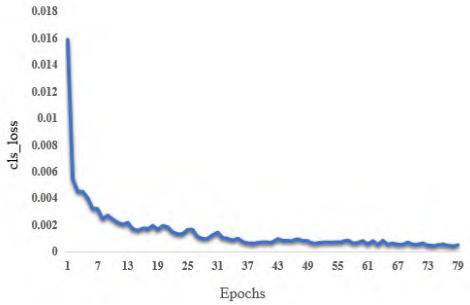
$$\upsilon = \frac{4}{\pi^2}(\arctan\frac{\omega^{gt}}{h^{gt}} - \arctan\frac{\omega}{h})^2$$

4.3 实验结果

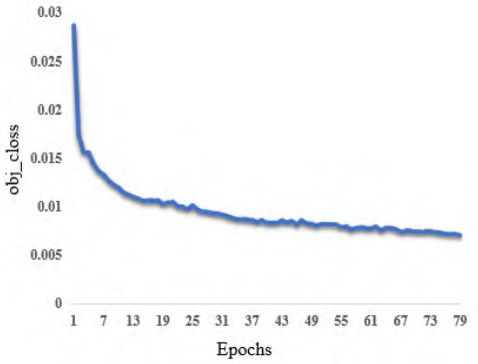
本文实验过程中损失函数变化如图 9 所示，从整体上来看，图像与表格的识别与提取效果一致。损失函数在前 10 次训练中迅速下降至 0.02 左右，随着训练轮次的增加，损失函数逐渐在 80 次训练中达到稳定，最后达到至 0.016 左右，模型收敛。分类损失函数（cls_loss）在训练中迅速下降至 0.004，由于表格为线框结构，且使用较为规范，干扰性因素较少，特征提取结果较为趋同，最终降至 0.00024，图像与表格的识别与分类效果良好。



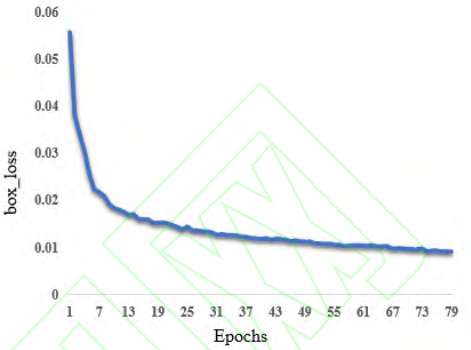
（a）loss 损失函数曲线



（b）cls_loss 损失函数曲线



（c）obj_loss 损失函数曲线

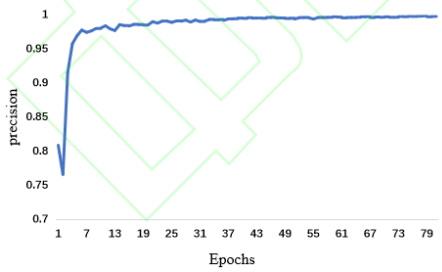


（d）box_loss 损失函数曲线

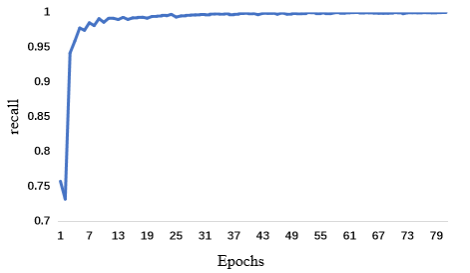
图 9　损失函数曲线

Fig.9 Loss Function Curve

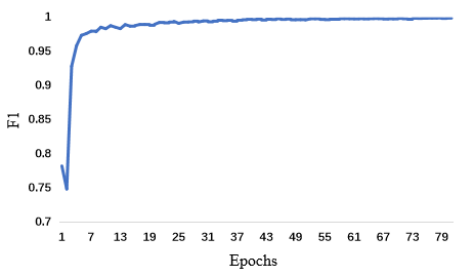
本文模型训练结果如图 10 所示，图 10（a）显示训练模型的精确度，在第四次训练后，模型的准确率已经超过 95%，随后逐渐趋于稳定，最大精确度为 99.84%；图 10（b）为召回率曲线，其最大值可达到 99.97%，可几乎全部识别标注图表；图 10（c）为 F1 曲线，最大值可取到 99.88%，模型表现良好；图 10（d）为 IOU 取值为 0.5 时，平均精度均值为 99.47%。综合来看，本文所构建的模型效果较好，针对学术期刊文献中的图表识别与提取有较好的效果。



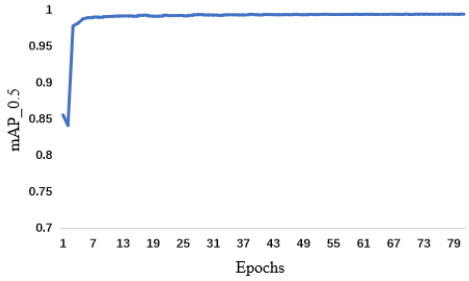
（a）精确率曲线



（b）召回率曲线



（c）F1 曲线



（d）mAP_0.5 曲线

图 10　模型训练结果

Fig.10 Model Training Effect

（1）消融实验

为了进一步验证各个优化模块的作用和有效性，本文进行了消融实验，实验结果如表 2 所示。传统视角下的 YOLOv5 整体效果较好，其原因主要是因为图表特征在学术期刊文献页面中的排版位置较为明显，在引入了 ECA 通道注意力模块后，F1 值提升了 1.5%，精确率和召回率也分别提升了 2.44%与 0.53%；在优化了 BiFPN 模块后，F1 值提升了 0.53%，精确率和召回率也有着少量的提升；最后将两者同时应用至模型中，整体效果较好，F1 值终达到了 99.88%，在深度学习中是一个较好的评价数值。

表 2　消融实验结果

Table2 Ablation Experimental Results

模型	ECA 模块	BiFPN 模块	mAP（%）	F1（%）	P（%）	R（%）
YOLOv5	×	×	95.28	97.89	96.58	99.23
优化 1	√	×	98.56	99.39	99.02	99.76
优化 2	×	√	96.98	98.42	97.5	99.35
YOLOv5-ECA-BiFPN	√	√	99.47	99.88	99.84	99.93

（2）基线模型

为综合验证本文优化算法的检测性能，本文选取目前最具有典型代表性的 One Stage 与 Two Stage 目标检测算法、YOLO 系列其他算法作为基线模型进行对比实验，具体如下：

Faster R-CNN：使用 VGG-16 网络结构，在 Fast R-CNN 基础上提出了 RPN 候选框生成算法。

SSD：在 VGG16 的基础上新增了卷积层来获得更多的特征图以用于检测。

YOLOv3：使用一个单独神经网络作用在图像上，将图像划分为多个区域并且预测边界框和每个区域的概率。

YOLOv4：引入 FPN+PAN 结构，回归框位置损失函数 CIOU_Loss，预测框筛选的 nms 变为 DIOU_nms。

实验结果如表 3 所示，实验采用了召回率，精确率、mAP 作为指标对不同的算法进行衡量。本文所构建的算法 mAP 值达到 99.47%，F1 值达到 99.88%，高于次优值 1.99%（SSD），相较于其他主流的目标检测算法有着较好的检测性能。

表 3　基线模型性能对比

Table3 Performance Comparison with Baseline Models

模型	mAP	F1	Precision	Recall
Faster R-CNN	92.35	93.24	92.83	93.65
SSD	95.28	97.89	96.58	99.23
YOLOv3	93.04	93.70	93.16	94.25
YOLOv4	93.25	94.29	93.87	94.72
YOLOv5-ECA-BiFPN	99.47	99.88	99.84	99.93

4.4 实例分析

为了进一步验证本文算法在实际应用场景的检测效果与泛化能力，本文选取现有文献中存在的六种特殊场景进行测试，确保算法在正常场景下的准确识别外，还能有效克服极端情况所带来的干扰；同时针对表格的提取进行检测，以确保算法对图表的区分能力。

（1）“噪音”图像干扰

“噪音”图像的干扰往往出现在首页与尾页，使得提取的效果中掺杂了众多无效信息，为了更好地展示本文提出的算法对噪音图像的“剔除”能力，实验将“噪音”图像添加至主体页面中，识别效果如图 11 所示。

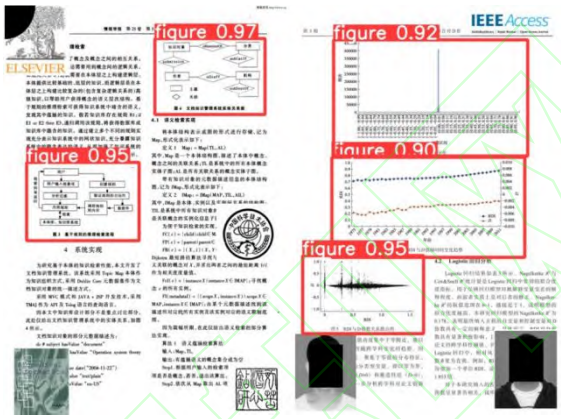


图 11　噪音图像干扰下的识别效果

Fig.11 Recognition Effect Under Noise Image Interference

（2）低像素图像

低像素图像广泛分布于 2010 年之前的学术期刊文献中，大部分为实物类图像与复杂类图像。本文提出的算法可以较好地识别低像素图像，识别结果如图 12 所示。

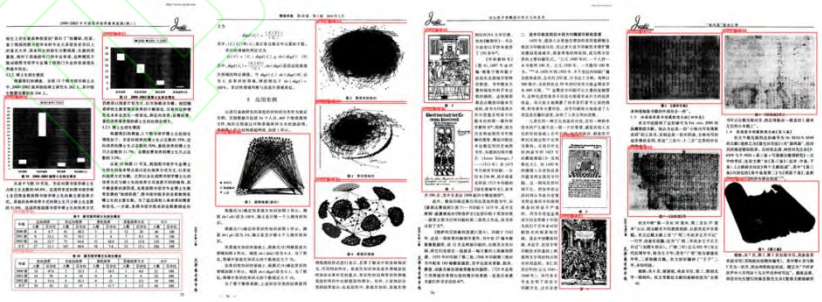


图 12　低像素图像识别效果

Fig.12 Recognition Effect of Low-Pixel Image

（3）图像分布不均匀

由于研究主题限制、出版社差异等因素，图像在排版布局上会出现分布不均匀的情况，通过分析均匀空间间隙的图像分离技术效果较差。本文提出的算法基于图像本体信息，可以较好地识别图像分布不均匀的情况，识别效果如图 13 所示。

图 13　不均匀分布图像识别效果

Fig.13 Recognition Effect of Uneven Distribution Image

（4）图像构成多样化

随着计算机与制图技术的发展，图像类型与其所包含的元素日益复杂，本文提出的算法可以排除颜色因素对识别结果的干扰（图 14(a)），文本构成图像与上下文本背景精准分离（图 14(b)），空间聚集程度低的图像不会过度分割（图 14(c)），同时还可以针对多类型的图表实现良好的提取效果，提取结果如图 15 所示。

(a) 颜色干扰　(b) 文本要素构成　(c) 聚集程度低

图 14　多样化图像识别效果

Fig.14 Recognition Effect of Diversified Image

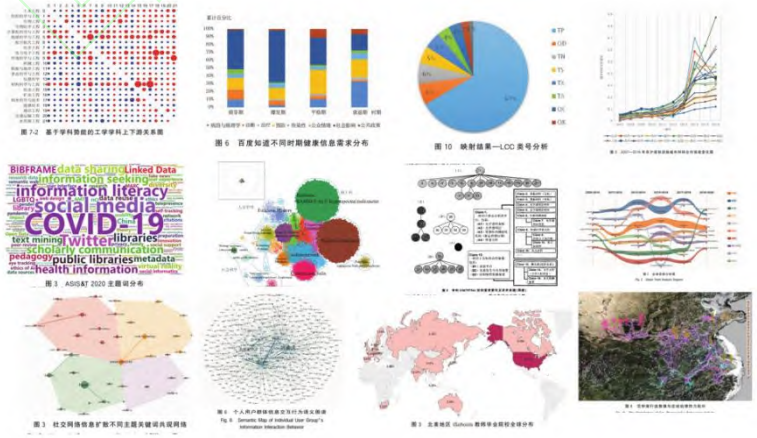


图 15　多类型图像提取结果

Fig.15 Extraction Results of Multi-Type Images

（5）标签信息与本体语义存在偏差

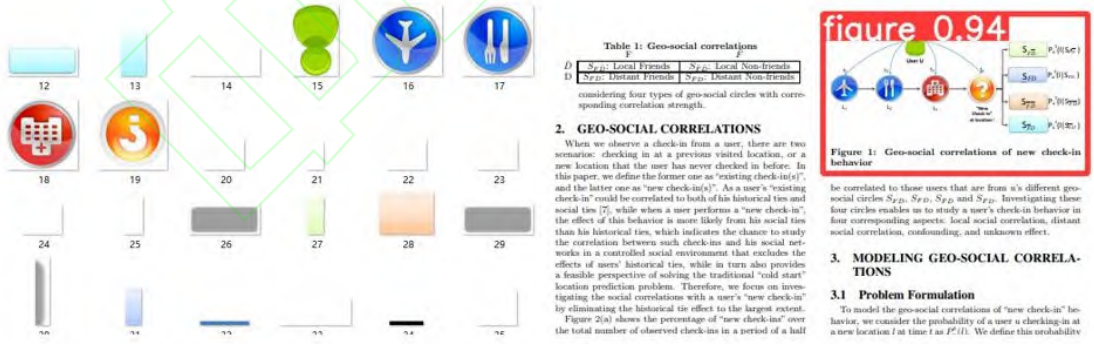
此类场景多存在于本体语义信息为表格时，标题信息缺失或以图像命名，导致无法使用标签信息遍历的方式进行表格识别与提取。本文提出的算法将语义信息与特征信息融合，无需依赖标题信息可实现图表的识别，识别结果如图 16 所示。

图 16　语义偏差图表提取效果

Fig.16 Recognition Effect of Semantic Deviation Image

（6）矢量图像构成复杂

本文选取 PDFfigures2[3]对矢量图像进行提取，该工具广泛应用至各类学术图表研究中。我们以文献（Modeling geo-social correlations for new check-ins on location-based social networks）为例，通过 PDFfigures2 提取矢量图像时，会将矢量元素构成的图像拆分错误，如图 17（a）所示，当采用本文提出的算法进行识别时，可以将图像信息全部提取，识别结果如图 17（b）所示。



(a) PDFfigures2识别效果　　(b) 本文算法识别效果

图 17　矢量图像识别效果对比

Fig.17 Vector Image Recognition Effect Comparison

（7）表格识别

在众多的方法中，大多是以图像识别与提取为主，基于表格的提取较少，本文提出的算法可以高效地对学术期刊文献中的表格进行识别，识别结果如图 18 所示。

图 18　表格识别效果

Fig.18Recognition Effect of Table

5　结语

本文构建的 YOLOv5-ECA-BiFPN 算法，能够对不同类型学术期刊文献中的图表进行识别与提取，从而解决特殊场景下图表提取效果差、泛化能力弱等问题，为后续学术可视化资源的研究与利用奠定基础。该算法在原有的 YOLOv5 目标识别基础之上，引入了 ECA 通道注意力模块与 BiFPN 模块，然后通过消融实验与对比实验，进一步证明该算法在学术图表识别过程中具有一定的优势，并设计了不同实验情境验证了其在多种情况下的识别能力和泛化能力。

与现有的研究相比，本文的创新点主要呈现在两个方面。首先，在研究方法上，本文首次将结构化 PDF 文献与非结构化 PDF 文献相结合，通过计算机视觉检测的方式在自建的数据集上进行图表识别的模型训练，利用深度学习的方式代替了人工规则的梳理，快速且高效地提取图像分布规则与排版方式；其次，根据图表在学术期刊文献中分布特点，结合 ECA 通道注意力模块与 BiFPN 模块优化模型，使得模型识别能力进一步提升，可以准确地识别学术期刊文献中的图表分布。但是本研究仍存在着一定的不足，模型训练样本需要依赖人工标注，样本量越大对于模型的识别检测越有帮助。后续笔者将扩大标注样本数据集，进一步提升多种情况下的图表识别能力。

参考文献：

[1]丁培.学术图表知识发现技术框架及研究进展[J].图书情报工作,2021,65(23):136-148.(Ding Pei.The Technical Framework and Research Progress of Knowledge Discovery in Academic Fiqures and Tables[J]Library And Information Service,,2021,65(23):136-148.)

[2]Liu Y, Si C, Jin K, et al. FCENet: An Instance Segmentation Model for Extracting Figures and Captions From Material Documents[J]. IEEE Access, 2020, 9: 551-564.

[3]Clark C,Divvala S. PDFFigures 2.0: Mining figures from research papers[C]// Acm/ieee-cs on Joint Conference on Digital Libraries. ACM, 2016.

[4]于丰畅,陆伟.一种学术文献图表位置标注数据集构建方法[J].数据分析与知识发现,2020,4(06):35-42.(Yu Fengchang,Lu Wei.Constructing Data Set for Location Annotations of Academic Literature Figures and Tables[J]. Data Analysis and Knowledge Discovery,2020,4(06):35-42.)

[5]Glyph&Cog.Xpdf[EB/OL].[2022-09-13].http://www.xp.dfreader.com.

[6]Choudhury S R ,Giles C L . An Architecture for Information Extraction from Figures in Digital Libraries[C]// International World Wide Web Conferences Steering Committee. International World Wide Web Conferences Steering Committee, 2015.
[7]Simon A, Pret J-C, Johnson A P. A Fast Algorithm for Bottom-Up Document Layout Analysis[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1997, 19(3): 273-277.
[8]ApacheSoftwareFoundation.ApachePDFBox[EB/OL].[2022-05-13].https://pdfbox.apache.org
[9]YUSUKE S. PDFMiner[EB/OL].[2022-09-13].https://github.com/euske/pdfminer
[10]T. Hassan. Object-level document analysis of pdf files. In Proceedings of the 9th ACM symposium on Document engineering, pages 47–55. ACM, 2009.
[11]于丰畅,程齐凯,陆伟.基于几何对象聚类的学术文献图表定位研究[J].数据分析与知识发现,2021,5(01):140-149.(Yu Fengchang,Cheng Qikai,Lu Wei.Locating Academic Literature Figures and Tables with Geometric Object Clustering[J]. Data Analysis and Knowledge Discovery,2021,5(01):140-149.)
[12]Praczyk, Piotr, Adam, et al. Automatic Extraction of Figures from Scientific Publications in High-Energy Physics.[J]. Information Technology & Libraries, 2013.
[13]Pengyuan, Li, Xiangying, et al. Figure and Caption Extraction from Biomedical Documents.[J]. Bioinformatics, 2019.
[14]Siegel N, Lourie N, Power R, et al. Extracting scientific figures with distantly supervised neural networks[C]//Proceedings of the 18th ACM/IEEE on joint conference on digital libraries. 2018: 223-232.
[15]Li P Y, Jiang X Y, Shatkay H. Figure and Caption Extraction from Biomedical Documents[J]. Bioinformatics, 2019, 35(21): 4381-4388.
[16]Chen K, Seuret M, Liwicki M, et al. Page Segmentation of Historical Document Images with Convolutional Autoencoders [C]//Proceedings of the International Conference on Document Analysis and Recognition. 2015.
[17]Amin A, Shiu R. Page Segmentation and Classification Utilizing Bottom-Up Approach[J]. International Journal of Image and Graphics, 2001, 1(2): 345-361.
[18]Gomez-Kramer, Petra, Heroux, et al. Texture feature benchmarking and evaluation for historical document image analysis[J]. International journal on document analysis and recognition, 2017.
[19]Ha J, Haralick R M, Phillips I T. Recursive X-Y Cut Using Bounding Boxes of Connected Components[C]// Proceedings of the 3rd International Conference on Document Analysis and Recognition. 1995: 952 .
[20]张建东,陈仕吉,徐小婷,左文革.基于词向量的 PDF 表格抽取研究[J].数据分析与知识发现,2021,5(08):34-44.
[21]Hassan T , Baumgartner R . Table Recognition and Understanding from PDF Files[C]// Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on. IEEE, 2007.
[22]唐锐,邓建新,叶志兴,张海平.PDF 文件的表格抽取研究综述[J].计算机应用与软件,2021,38(07):1-7+22.
[23]Ultralytics.YOLOV5[CP/OL].[2022-11-12]. https: //github.com/ultralytics/yolov5.
[24]Lin T Y, Dollár P, Girshick R, et al. Feature pyramid networks for object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 2117-2125.
[25]Liu S, Qi L, Qin H, et al. Path aggregation network for instance segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 8759-8768.
[26]Wang Q, Wu B, Zhu P, et al. Supplementary material for‘ECA-Net: Efficient channel attention for deep convolutional neural networks[C]//Proceedings of the 2020 IEEE/CVF Conference on
[27]Tan M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790

作者贡献声明：

李英群：论文起草；

李亚菲：提出研究思路，设计研究方案，修改论文；

裴雷：完善研究方案，修改论文；

胡志伟，宋宁远：论文最终版本修订。

利益冲突声明：

所有作者声明不存在利益冲突关系。