

文章编号: 2095-2163(2024)01-0194-06

中图分类号: TP391.4

文献标志码: A

基于 YOLOv7 改进模型的苹果质量识别算法研究

任晓东

(天津职业技术师范大学 电子工程学院, 天津 300222)

摘要: 针对苹果质量分拣的识别, 提出一种基于 YOLOv7 改进模型的苹果质量识别算法。通过引入 SimAM 无参数注意力机制模块, 以提高网络的特征提取能力, 改善苹果质量识别的精度。结果表明, 本研究提出的 YOLOv7-SimAM 模型平均精度均值 mAP 达到了 88.2%, 比原模型提升了 2.2%, 该算法可以快速准确的识别苹果的质量, 提高水果加工的工作效率。

关键词: YOLOv7; 苹果质量识别; SimAM; 无参数注意力机制

Research on apple quality recognition algorithm based on YOLOv7 improved model

REN Xiaodong

(School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China)

Abstract: Aiming at the identification of apple quality sorting, an apple quality recognition algorithm based on the improved YOLOv7 model is proposed. By introducing the SimAM parameter-free attention mechanism module, the feature extraction ability of the network is improved and the accuracy of apple quality recognition is improved. The results show that the average precision mean mAP of YOLOv7-SimAM model proposed in this study reaches 88.2%, which is 2.2% higher than that of the original model. The algorithm can quickly and accurately identify the quality of apples and improve the efficiency of fruit processing.

Key words: YOLOv7; apple quality recognition; SimAM; parameter-free attention mechanism

0 引言

中国是苹果最大的生产国之一, 但是目前大部分的苹果加工厂分拣都是人为识别, 存在检测率低下的问题, 而手动分拣一不小心会碰坏水果, 以至于引起水果后期储存时间的大大减少, 使水果加工厂的经济效益受到一定的影响。

随着深度学习的发展, 面向水果果实的质量检测得到了广泛的研究。Kuang H 等人^[1]提出了一种基于图像区域选择和改进目标建议的多类别水果检测方法(MCRD), 该方法利用了有效的图像区域选择和改进目标建议。Jidong L 等人^[2]所做的研究主要是在自然环境下识别封闭苹果, 对于重叠状态下的苹果, 则通过在边缘图像中加入分割的二值图像的方法进行处理, 分离出苹果果实之间的重叠和粘连部分。项辉宇等^[3-4]使用基于图像处理方法, 对苹果的大小、颜色和缺陷进行了检测, 该方法主要是基于水果形状、边缘、纹理等普通特征来定位缺陷目标。高辉等^[5]采用自动亮度校正技术与加权矢量机相连接, 进一步提高了水果缺陷检测准确率和

检测速度, 但该方法需要进行光照系统的配置, 成本较大。薛勇等^[6]采用 Google Net 深度迁移模型对苹果缺陷进行检测, 获得了较好的泛化能力与鲁棒性, 但该卷积网络模型复杂、参数多且检测效率低。夏雪等^[7]构建了用于苹果检测的轻量级无锚点深度学习网络模型 M-Center Net, 该方法通过改进轻量级 MobileNetV3 网络, 结合关键点预测的目标检测网络(Center Net), 为非结构环境下果园作业平台的轻量化果实目标检测模型研究提供了新的思路。

综上所述, 本文拟构建一阶段目标检测算法 YOLOv7^[8], 结合无参注意力模块(SimAM^[9])模型解决苹果质量识别问题。

1 YOLOv7 目标检测算法

1.1 YOLOv7 结构

YOLO 算法是基于神经网络对目标对象进行识别和定位, 不仅作为一阶段目标检测的典型代表, 而且其运行速度很快, 能够运用到实时识别检测系统中。而 YOLOv7 作为如今 YOLO 系列中最出色的算法, 准确率和速度均优于之前的 YOLO 算法,

作者简介: 任晓东(1997-), 男, 硕士研究生, 主要研究方向: 人工智能、深度学习算法。

收稿日期: 2022-12-13

YOLOv7 整体结构如图 1 所示。在输入部分将图片调整为 640×640 后, 输入到 backbone 网络中; 经过

head 网络输出 3 层不同大小的特征图谱, 再经由 Rep 和 conv 输出预测结果。

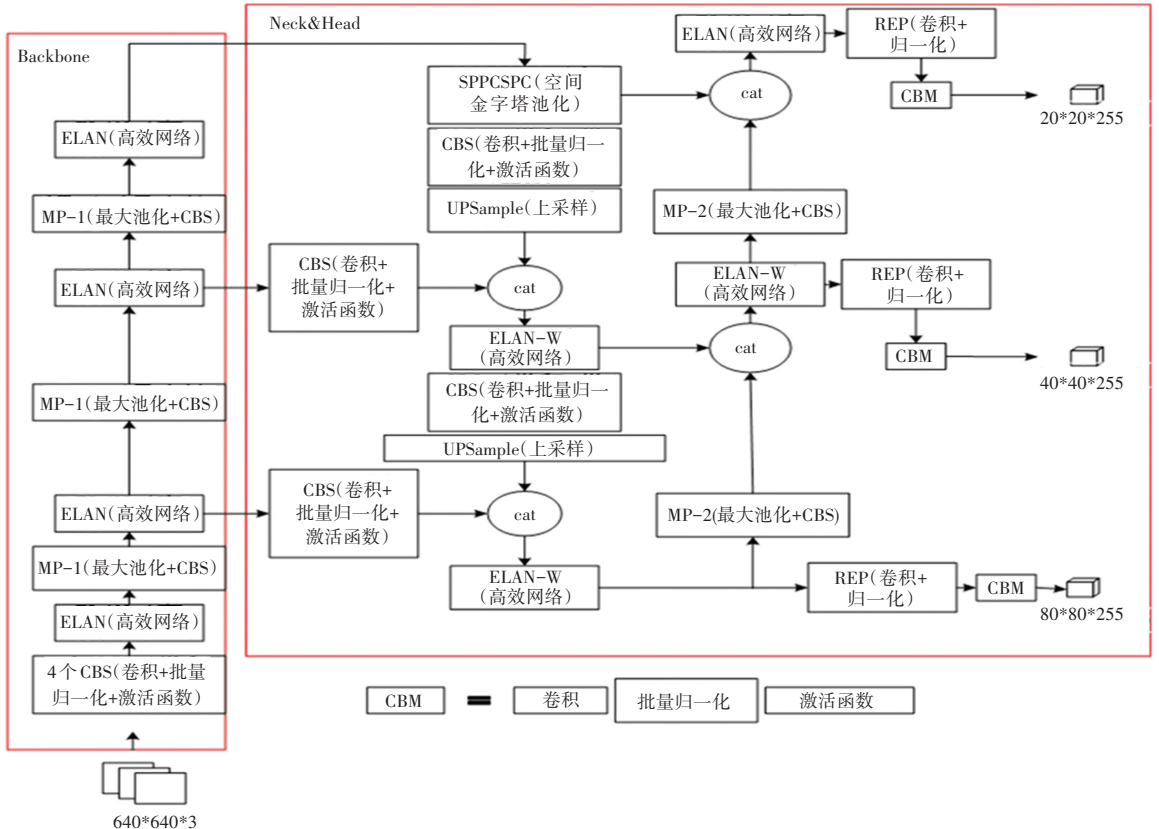


图 1 YOLOv7 整体结构

Fig. 1 Overall structure of YOLOv7

1.2 ELAN 模块

ELAN^[10] 是 YOLOv7 模型中提出的一个新网络结构, 其主要以高效为主, 可以通过控制最长和最短的梯度路径, 使得网络能够学习到更多的特征, 而且拥有更强的鲁棒性。如图 2 所示, ELAN 模块结构

有两个分支。其中一个分支仅经过一个 1×1 的卷积做通道数的变化; 而另一个分支先经过一个 1×1 的卷积模块来做通道变化, 再经过 4 个 3×3 的卷积模块来做特征提取, 最后将 3 个特征叠加在一起, 再经过一个 CBS 模块输出即为需求的通道。

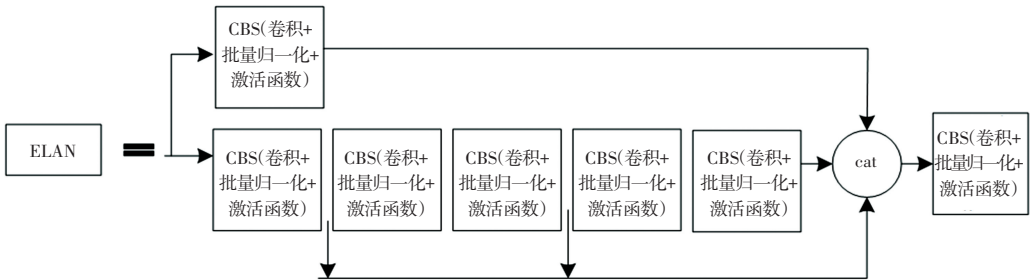


图 2 ELAN 模块

Fig. 2 ELAN module

1.3 CBS 模块

CBS 模块由一个卷积层 (Conv)、一个 BN 层和 Silu 层组成, 其结构如图 3 所示。CBS 模块有 3 种形式, 但其作用各不相同, 卷积核和步长也不同。一种是 1×1 的卷积, 步长为 1, 主要作用是改变通道数; 第二种是 3×3 的卷积, 步长为 1, 主要作用是进

行特征提取; 最后一种是 3×3 的卷积, 步长为 2, 主要用作下采样。



图 3 CBS 模块

Fig. 3 CBS module

1.4 SPPCSPC 模块

SPP 模块的作用是可以增大感受野,让算法适应不同的分辨率图像。如图 4 所示,最大池化有不同的尺度,所以就获得了不同的感受野,用来区别于

大目标和小目标。CSP 模块首先将特征分成两部分,一部分进行常规的处理,另一部分进行 SPP 结构的处理,最后将两部分合并在一块,因此减少了一半的计算量,从而使得速度得以提高。

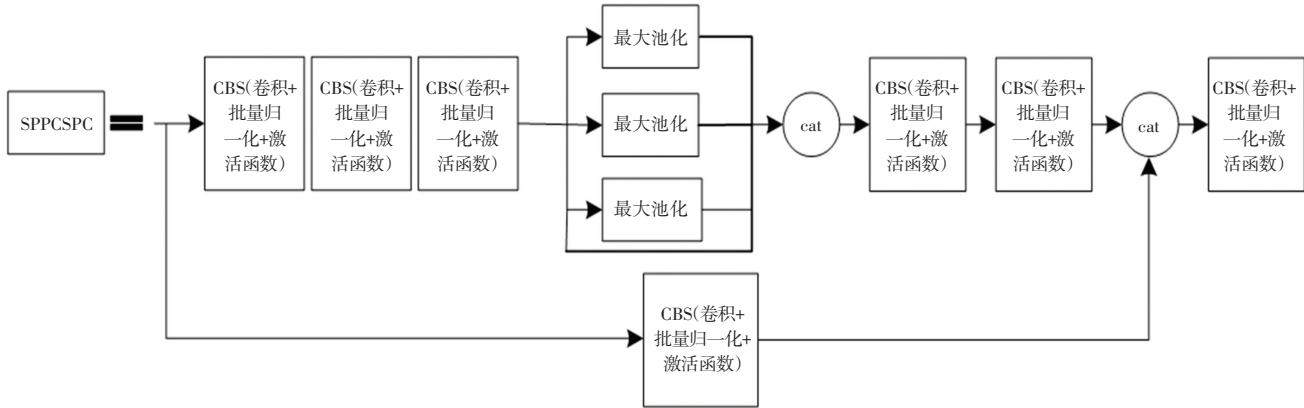


图 4 SPPCSPC 模块

Fig. 4 SPPCSPC module

1.5 REP 模块

如图 5 所示,REP 模块分为训练模块和推理模块。训练模块有 3 个分支,第一个分支是 3×3 的卷积,第二个分支为 1×1 的卷积用来平滑特征,最后的分支为恒等映射模型卷积重参数化^[11]时,把 1×1 的卷积转化成 3×3 的卷积,而后恒等映射也转为 3×3 的卷积,再对其进行矩阵加法,最后将其权重进行加和,就获得了一个 3×3 的卷积;而推理模块则仅含有一个 3×3 的卷积,步长为 1,该卷积是由重参数化转化而来。

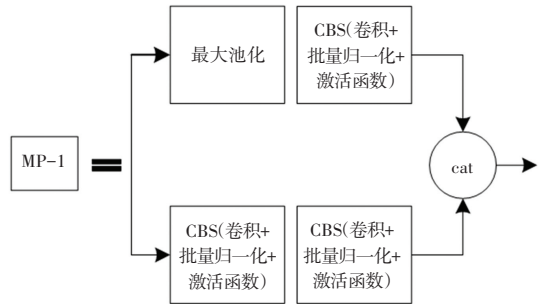


图 6 MP 模块

Fig. 6 MP module

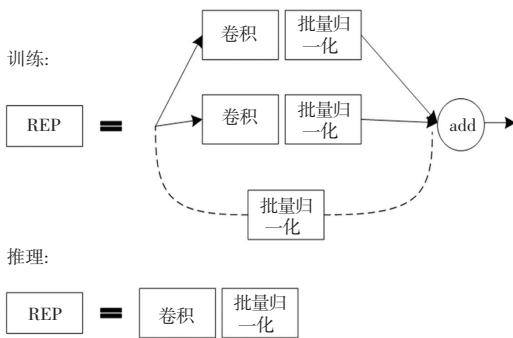


图 5 REP 模块

Fig. 5 REP module

1.6 MP 模块

如图 6 所示,MP 模块的作用是进行下采样。从图中得知,第一个分支首先经过最大池化,通道数的改变是通过一个 1×1 卷积来实现的;第二个分支首先经过一个 1×1 卷积完成对通道数变化的任务,再历经一个下采样的 3×3 卷积、步长为 2 的卷积块,最后将第一和第二分支的结果相加,就得到了超级下采样的结果。

2 YOLOV7 网络改进

2.1 SimAM 无参数注意力

SimAM^[9]与现有的通道、空域注意力模块是有所区别的,该模块无需额外的参数为特征图来推导出 3D 注意力权重。SimAM 基于一些闻名的神经科学理论,通过提出优化一个能量函数,以便于找到每一个神经元的重要性。SimAM 的优点是:按照所定义的能量函数的解来挑选大多数算子,避免了在结构调整上消耗的精力。

评估每个神经元的重要性,目的是为了更好的实现注意力。在神经科学领域中,拥有大量信息的神经元往往显示出和周围神经元不同的放电模式,况且激活神经元经常会抑制周围神经元,即为空域抑制。对于拥有空域抑制效应的神经元给予其更高的重要性。度量神经元之间的线性可分性是一种最简单的寻找神经元的方法。因此,定义如下的能量函数如式(1)所示:

$$e_i(\omega_i, b_i, y, x_i) = (y_i - \hat{i})^2 + \frac{1}{M-1} \sum_{i=1}^{M-1} (y_o - \hat{x}_i)^2 \quad (1)$$

最小化上述公式等价于训练同一通道内的神经元 t 与其他神经元之间的线性可分性。为了简单起见,采用二值标签,并且添加正则项,最后的能量函数定义如式(2)所示:

$$e_i(\omega_i, b_i, y, x_i) = \frac{1}{M-1} \sum_{i=1}^{M-1} (-1 - (\omega_i x_i + b_i))^2 + (-1 - (\omega_i t + b_i))^2 + \lambda \omega_i^2 \quad (2)$$

式中:

$$\omega_i = -\frac{2(t - \mu_i)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda} \quad (3)$$

$$b_i = -\frac{1}{2}(t + \mu_i) \omega_i \quad (4)$$

由于每个通道上全部神经元都遵循一样的分布,所以先对输入特征在高度和宽度两个维度上计算均值和方差,避免重复计算,如式(5)所示:

$$e_i^* = \frac{4(\hat{\sigma}^2 + \lambda)}{(t - \mu_i)^2 + 2\sigma_i^2 + 2\lambda} \quad (5)$$

整个过程可以用式(6)表示为

$$\tilde{X} = \text{sigmoid}\left(\frac{1}{E}\right) \odot X \quad (6)$$

最后,SimAM 注意力机制直接推断三维权重如图7所示,从图中看出,相同的颜色表示为每个通道、空间位置或该特征上每个点使用的单个标量。而且该无参数注意力机制更灵活,也保持着轻量级。

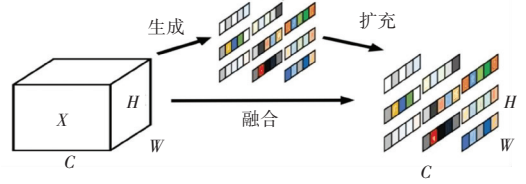


图 7 全三维权重注意力

Fig. 7 Full 3-D weights for attention

2.2 网络改进

本文以 YOLOv7 网络模型为基础进行苹果质量的识别,为了提高苹果质量的识别精度和检测的准确率,在网络结构 head 部分的卷积层添加无参数注意力机制 SimAM,添加注意力机制后的 YOLOv7 模型称为 YOLOv7-SimAM。

提出 YOLOv7-SimAM 模型与 YOLOv7 原模型相比较,对于苹果质量识别有较好的检测精度。该模型的网络架构如图 8 所示。

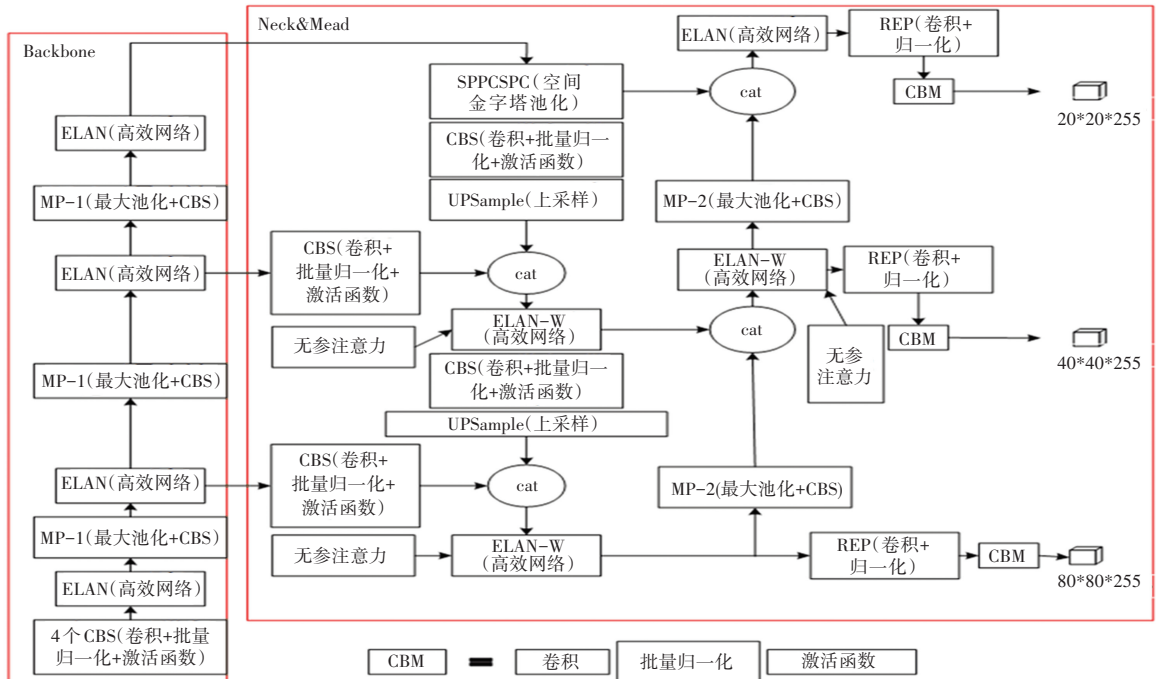


图 8 改进的 YOLOv7 模型结构

Fig. 8 Improved YOLOv7 model structure

3 实验与分析

3.1 实验环境

本文使用 Pytorch 框架构建网络模型,程序在 Windows11 系统下运行,训练模型的处理器为 Intel (R)Core(TM) i7-12700K 3.61 GHz,内存 16 GB 和显存 11 GB 的 NVIDIA GeForce RTX3060, CUDA 版本为 10.0。

3.2 评估指标

评价一个模型性能的好坏,关键在于评估指标的真实性。在二分类问题中,根据样本的真实类别和模型预测类别的组合从而将其分为:预测为正的样本(TP)、预测为负的正样本(FN)、预测为正的负样本(FP)、预测为负的负样本(TN)。本文采用精准率(Precision)、召回率(Recall)和平均精度均值(mAP)作为评估指标。精准率 P 越高则表示预测结果中正确的结果样本占比越高,误检越低;召回率 R 越高则表示预测结果中正样本被正确检测出来的越多,漏检越低;平均精度均值 mAP 越高,目标检测模型的各个类别的平均检测效果越好。计算公式如式(7)~式(9)所示:

$$P = \frac{TP}{TP + FP} \times 100\% \quad (7)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (8)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad (9)$$

其中, AP 的计算公式如式(10)所示:

$$AP = \int_0^1 p(r) dr \quad (10)$$

3.3 数据集

本文使用 Kaggle 网站中的 Fruits fresh and rotten for classification 数据集,以及 Fruits360 数据集的部分图片作为苹果质量识别数据集。为了提高模型的泛化能力,本文对该数据集采取数据增强技术,如:平移、翻转、旋转、添加噪声、改变图像亮度等方法,从而提高模型的鲁棒性。针对该数据集,使用 Labellmg 标注工具进行标注,输出生成 YOLO 框架的 TXT 类型的标签文件。数据集按照 7:3 的比例划分训练集和验证集,分别得到训练集 518 张图片,验证集 222 张图片。

3.4 模块训练及性能对比

本文选择 YOLOv7 网络作为苹果质量识别算法的基础网络模型,在网络模型训练前,设置迭代批量

大小为 4,迭代次数设置为 200 次,初始学习率设置为 0.001,然后分别在原模型和改进的模型上对苹果数据集进行训练,并对精度等指标进行分析,结果见表 1。表中数据说明,改进的 YOLOv7-SimAM 模型相比 YOLOv7 模型的 mAP 提升了 2.2%。

表 1 原模型与改进模型性能对比

Table 1 Performance comparison between the original model and the improved model %

网络模型	准确率	召回率	平均精度
YOLOv7	85.1	86.5	86
YOLOv7-SimAM	82.9	92.9	88.2

本文所提出的模型与原模型 PR 曲线如图 9 所示。从图中可以看出,改进的模型检测识别质量优质的苹果为 85.7%,腐烂、虫咬以及磕碰的质量劣质苹果为 90.8%,均比原模型识别的准确率高。结合表 1 结果可以看出,YOLOv7-SimAM 模型比 YOLOv7 模型的苹果质量识别检测能力更强,精度更高。因此,改进后的模型在苹果质量检测上获得了较好的结果。

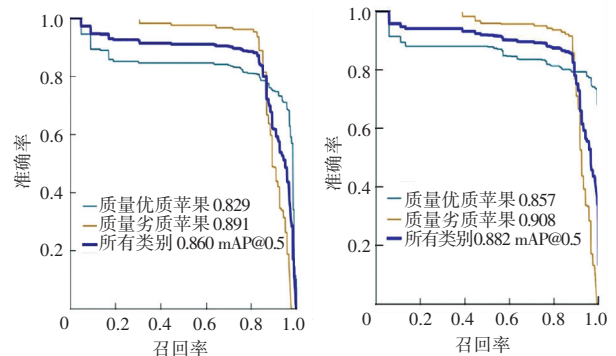


图 9 模型改进前后 PR 曲线

Fig. 9 PR curves before and after the improved model

4 结束语

本文针对苹果质量识别的研究,提出了一种基于 YOLOv7 改进模型的苹果检测算法。改进模型将无参数注意力机制 SimAM 和原始的 YOLOv7 相结合,从而在模型网络中 mAP 从原来的 86%到现在的 88.2%,提高了 2.2%;召回率从 86.5%提高到现在的 92.9%。实验结果表明,加入 SimAM 无参注意力的 YOLOv7 模型在准确率方面有所提高,mAP 也得到了提升。因此将此研究算法应用到水果加工厂的苹果识别中,既能提高工厂的识别效率,也能减轻工人的劳动力强度。

参考文献

- [1] KUANG H, LIU C, et al. Multi-class fruit detection based on image region selection and improved object proposals [J]. *Neurocomputing*, 2018, 283: 241–255.
- [2] JIDONG L, DE-AN Z, WEI J, et al. Recognition of apple fruit in natural environment[J]. *Optik*, 2016, 127(3): 1354–1362.
- [3] 项辉宇, 薛真, 冷崇杰, 等. 基于 Halcon 的苹果品质视觉检测试验研究[J]. *食品与机械*, 2016, 32(10): 123–126.
- [4] 李红, 杨颖辉. 基于混沌多宇宙算法的苹果表面缺陷检测研究[J]. *江苏农业科学*, 2017, 45(15): 202–205.
- [5] 高辉, 马国峰, 刘伟杰. 基于机器视觉的苹果缺陷快速检测方法研究[J]. *食品与机械*, 2020, 36(10): 125–129, 148.
- [6] 薛勇, 王立扬, 张瑜, 等. 基于 Google Net 深度迁移学习的苹果缺陷检测方法[J]. *农业机械学报*, 2020, 51(7): 30–35.
- [7] 夏雪, 孙琦鑫, 侍啸, 等. 基于轻量级无锚点深度卷积神经网络

- 的树上苹果检测模型[J]. *智慧农业(中英文)*, 2020, 2(1): 99–110.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[J]. *arXiv preprint arXiv:2207.02696*, 2022.
- [9] YANG L, ZHANG R Y, LI L, et al. Simam: A simple, parameter-free attention module for convolutional neural networks [C]//International Conference on Machine Learning. PMLR, 2021: 11863–11874.
- [10] ZHANG X, ZENG H, GUO S, et al. Efficient long-range attention network for image super-resolution [C]//European Conference on Computer Vision. Cham: Springer Nature Switzerland, 2022: 649–667.
- [11] DING X, ZHANG X, MA N, et al. Repvgg: Making vgg-style convnets great again[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13733–13742.

(上接第 193 页)

表 3 多模态情感分析结果

Table 3 Results of multimodal emotion analysis

融合方式	模型	准确率	宏平均 F1	加权平均 F1
直接拼接	CNN	0.690	0.682	0.703
	BiLSTM	0.700	0.685	0.711
	BiGRU	0.685	0.663	0.672
	CNN+attention	0.696	0.696	0.694
	BiLSTM+attention	0.705	0.706	0.711
	BiGRU+attention	0.715	0.704	0.723
线性融合	CNN	0.701	0.705	0.703
	BiLSTM	0.696	0.692	0.696
	BiGRU	0.712	0.715	0.710
	CNN+attention	0.697	0.692	0.698
	BiLSTM+attention	0.707	0.708	0.702
	BiGRU+attention	0.736	0.737	0.735

由直接拼接和线性融合方式的多模态网络舆情情感分析结果可以看出,相对于直接拼接特征融合方式,线性融合方式的特征融合所构建的模型各评价指标值更高。本文的融入注意力机制的双向门控神经网络模型准确率为 0.736,宏平均 F1 值为 0.737,加权平均 F1 值为 0.735,相比其他基线模型,效果更佳,且准确率、宏平均 F1 和加权平均 F1 的值很相近,说明本文提出的多模态网络舆情情感分析框架对于舆情事件的正面、中立和负面的情感识别能力相当,并没有因为数据分布的不平衡而偏向识别哪一类情感,具有可扩展性,网络舆情情感分析的准确度高。

3 结束语

本文提出了一种基于 BiGRU 模型的多模态网

络舆情情感分析框架,加入网络舆情图片信息对纯文本信息进行补充,避免了单模态情感识别对情感极性的误判,提升了准确率。该模型首先对数据进行数据增强,利用 Word2vec 模型提取文本特征,利用 CNN 提取图像特征,使用线性融合对“图文对”进行特征层融合,减少信息冗余和有效抑制噪声,最后将纯文本特征与“图文对”特征一起输入到融合注意力机制的双向门控循环神经网络中,进行多模态情感分析。以“李宁天价鞋”真实数据作为研究对象,对上述模型进行了实证检验,并与不同的基线模型进行对比分析。实验结果表明,本文提出的多模态网络舆情情感分析框架对小样本的网络舆情事件情感分析具有较好地分析效果。

参考文献

- [1] 李光敏, 张行文, 张磊, 等. 面向网络舆情的评论文本情感分析研究[J]. *情报杂志*, 2014, 33(5): 157–160.
- [2] 宛艳萍, 谷佳真, 张芳. 融合改进 Stacking 与规则的文本情感分析[J]. *小型微型计算机系统*, 2021, 42(7): 1389–1395.
- [3] 刘颖, 王哲, 房杰, 等. 基于图文融合的多模态舆情分析[J]. *计算机科学与探索*, 2022, 16(6): 1260–1278.
- [4] 凌海彬, 缪裕青, 张万桢, 等. 多特征融合的图文微博情感分析[J]. *计算机应用研究*, 2020, 37(7): 1935–1939, 1951.
- [5] DU Y, LIU Y, PENG Z, et al. Gated attention fusion network for multimodal sentiment classification [J]. *Knowledge - Based Systems*, 2022, 240: 108107.
- [6] 陈小敏, 许华虎, 方迪恺, 等. 基于注意力机制的 TAI-CNN 图文跨模态情感分类方法[J]. *计算机应用与软件*, 2021, 38(5): 196–200, 293.
- [7] ZHANG D, XU H, SU Z, et al. Chinese comments sentiment classification based on word2vec and SVMperf[J]. *Expert Systems with Applications*, 2015, 42(4): 1857–1863.