

DOI: 10.13652/j.spjx.1003.5788.2024.80643

# 基于图像采集优化识别的白酒酒花分类方法

赵 谦<sup>1</sup> 孙 炎<sup>1,2</sup>

(1. 西安科技大学通信与信息工程学院, 陕西 西安 710054;

2. 山西杏花村汾酒集团有限责任公司, 山西 吕梁 032205)

**摘要:** [目的] 实现白酒酒段的自动化识别与分类, 解决白酒传统“看花摘酒”出品波动以及利用深度学习的酒花分类方法在精度、实时和普适性平衡中的问题。[方法] 提出一种基于图像采集优化识别的白酒酒花自动分类方法。通过自建平台采集酒花图像, 并利用 ENet 进行预处理以提高数据质量, 使用 Vision Transformer (ViT) 和 ConvNeXt 模型对酒花图像进行分类。[结果] 试验方法提高了白酒摘酒过程的自动化水平和精确度, 在保证实时性的同时, 分类准确率为 99.4%。[结论] 该方法有效优化了传统白酒摘酒工艺, 可以快速准确地实现酒花实时检测分类。

**关键词:** 白酒酒花; 机器视觉; 图像分类; 视觉检测; 实时分类; ConvNeXt; ENet

## Foam classification method of Chinese spirits based on image acquisition optimization recognition

ZHAO Qian<sup>1</sup> SUN Yan<sup>1,2</sup>

(1. College of Communication and Information Engineering, Xi'an University of Science and Technology, Xi'an, Shaanxi 710054, China; 2. Shanxi Xinghua Village Fenjiu Group Co., Ltd., Luliang, Shanxi 032205, China)

**Abstract:** [Objective] This paper aims to achieve automatic identification and classification of Chinese spirits, solve the production fluctuation of traditional "alcoholic strength determination based on foam watching" for Chinese spirits, and balance the accuracy, real-time performance, and universality of existing deep learning-based methods for foam classification of Chinese spirits. [Methods] An automatic foam classification method of Chinese spirit based on image acquisition optimization recognition was proposed. The foam images were collected through a self-built platform, and the data quality was improved by preprocessing via ENet. The foam images were classified by using the Vision Transformer (ViT) and ConvNeXt models. [Results] This method improved the automation level and accuracy of alcoholic strength determination for Chinese spirits and achieved a classification accuracy of 99.4% while ensuring real-time performance. [Conclusion] This method effectively optimizes the traditional alcoholic strength determination technology for Chinese spirits, enabling rapid and accurate real-time detection and classification of foams.

**Keywords:** foam of Chinese spirit; machine vision; image classification; visual detection; real-time classification; ConvNeXt; ENet

在白酒的生产过程中, 通常采用蒸馏取酒的方法进行酒液的提取。白酒蒸馏取酒过程中, 酒精度会不断降低, 同时白酒中的风味物质如酸类物质、酯类物质、醛类物质及杂醇类物质(甲醇等)会受到自身挥发性及其在酒精中溶解度的影响<sup>[1]</sup>。挥发性越高, 白酒中风味物质在乙醇中的溶解度越大, 会更快地被蒸馏而产生酒体<sup>[2]</sup>。因此, 不同时刻的酒液中的酒精度以及微量物质含量存在差异, 这直接导致了酒液的品质存在差异, 从而形成多个

不同的酒段。以某清香型白酒为例, 该清香型白酒在蒸馏出酒过程中, 酒液分为头酒、大清花、小清花、云花、二花、水花、油花 7 个酒段, 按时间顺序依次馏出<sup>[3]</sup>。

在摘酒过程中, 大部分企业和酒厂采用传统“看花摘酒”, 即有经验的师傅通过观察酒花(白酒酒液冲击、震荡产生的气泡)的形态、大小等来判断白酒分段<sup>[4]</sup>。这种传统工艺存在较多问题, 如个人经验依赖性强、效率低、工作强度大; 对摘酒工人的健康也造成影响等。而智能摘

基金项目: 陕西省教育厅服务地方企业项目(编号: 22JC050)

通信作者: 孙炎(1999—), 男, 西安科技大学在读硕士研究生。E-mail: 1329811605@qq.com

收稿日期: 2024-07-08 改回日期: 2024-11-27

酒在解决以上问题之余,还可以通过监测酒花气泡类型时间反向监督其他环节,作为生产的评判标准,起到反向监督的作用。

目前,有关智能摘酒方向的研究主要有基于分析仪器和计算机视觉两个方向。基于分析仪器的摘酒方法采用分析仪器如光谱仪等对酒体中的化学成分进行检测和分析,并以此作为判断分段的依据<sup>[5]</sup>。但该方法设备成本高且检测速度较慢,很难做到实时摘酒的要求<sup>[6]</sup>。基于计算机视觉的摘酒方法主要是通过对酒花图像的采集、优化和判断,从视觉特征模拟“看花摘酒”,完成对白酒酒段的判断过程。杨静娴等<sup>[7]</sup>结合局部二值模式(LBP)与灰度共生矩阵(GLCM)提取酒花纹理特征,通过支持向量机(SVM)得到分类结果。潘斌等<sup>[8]</sup>对 VGG16 卷积神经网络进行了改进和应用,通过高斯滤波后的酒花图像进行分类,但数据集较少且处理速度及精度均有待提升。余锴鑫<sup>[9]</sup>使用图像分类算法进行自动化的摘酒研究,但其前景分割算法使用椭圆曲线检测,方法复杂,仅前景分离部分在服务器上耗时 4.2 s,无法实时检测。田子宸<sup>[10]</sup>在余锴鑫<sup>[9]</sup>的基础上改进了前景分离算法,配合其设计的采集平台在速度上有较大突破,但需要使用其设计的平台,前景检测速度提升主要通过平台设计上的排除干扰,无法适用于其他平台和场景,普适性较差。计算机视觉摘酒的可解释性不如基于分析仪器的方法,但设备成本适中,安装维护相对便捷,精度与速度均能达到实际生产对自动化摘酒的需求。

研究拟针对以上方法的不足,提出一套基于白酒酒花视觉采集优化识别的酒花图像分类方法,通过自建平台,处理自建数据集的图像,并利用 ENet 进行预处理以提高数据质量,使用 Vision Transformer(ViT)和 ConvNeXt 模型对酒花图像进行分类,以期实现对酒花图像的实时分类,从而实现智能摘酒。

## 1 酒花视觉特征分析及数据集构建

### 1.1 酒花视觉特征分析

不同成分的酒体在相同的温度和压强环境下酒花的气泡形态不同<sup>[11]</sup>,以某清香型白酒为例,该清香型白酒酒花气泡特征及对应酒度关系见表 1。

表 1 酒花气泡特征及对应酒度关系

Table 1 Characteristics of foams and corresponding alcoholic strength

酒花名称	特征	酒精度/%Vol
大清花	气泡大如黄豆,均匀透亮	70~80
小清花	气泡小如绿豆,细密均匀	60~70
云花	酒花混乱,大小不一	40~50
二花	颜色无光、泡沫粘连	30左右
水花	泡沫大而无光	<20

七段酒中的酒头和酒尾阶段不根据酒花形态判断,因此不列入表 1。表 1 中 5 个阶段白酒酒花气泡形态如图 1 所示。

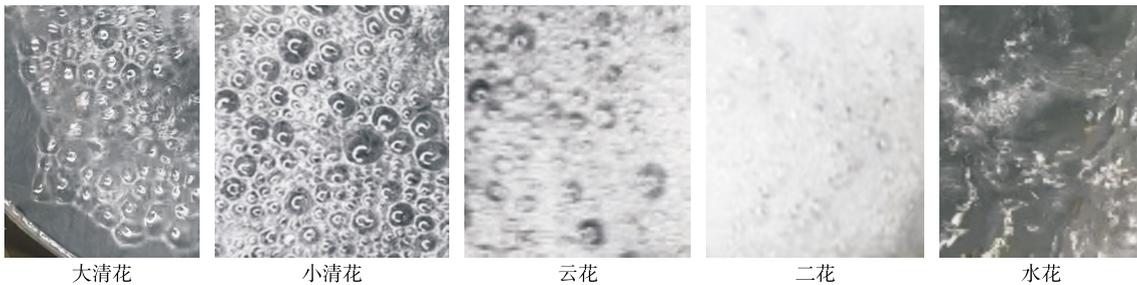


图 1 不同阶段酒花气泡形态

Figure 1 Foam morphologies at different stages

白酒在不同酒段较为明显的视觉特征差距是受多方因素影响的,如温度、压强及酒液成分等。在酒液成分中,对酒花视觉特征产生最主要影响的是乙醇含量。30 °C下,乙醇水溶液的表面张力与乙醇浓度  $c$  之间的拟合关系为<sup>[12]</sup>

$$\sigma = \frac{0.2243}{3.182 + c^{0.6899}} \quad (1)$$

随着乙醇含量的减小,乙醇水溶液的表面张力逐渐增大。整个蒸馏取酒过程中,酒液中乙醇含量随时间推移不断下降。因此,溶液的表面张力随之升高,导致泡沫

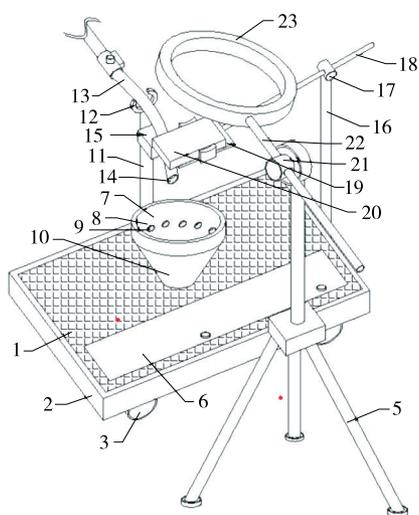
不易成形且稳定性差。故大清花、小清花及云花阶段泡沫较多,到了水花阶段,酒体酒精度下降较快,几乎没有成形的气泡。

此外,随着酒精度的不断降低,白酒中的风味物质的溶解度也会对酒花气泡的起泡和消泡产生影响<sup>[13]</sup>。大清花阶段酒精度高,有机酸、乳酸乙酯等易溶于水的风味物质含量低<sup>[14]</sup>,起泡作用不易产生,只有当气泡兼并到一定程度,起泡剂分子吸附量才足以维持泡沫的稳定,所以造成了大清花阶段气泡大如黄豆的特征<sup>[15]</sup>;小清花部分酒精度开始降低,易溶于水的风味物质含量增多<sup>[16]</sup>,起泡作

用更加明显,所以小清花阶段的特点为小如绿豆,细密均匀;而到了水花、油花阶段,酒精度快速降低,易溶于水的风味物质溶解度过高,反而降低了气泡的稳定性,酒液中有消泡作用的高级醇<sup>[17]</sup>、酯类物质开始起主导作用,使得水花、油花阶段几乎没有成型的气泡<sup>[18]</sup>。

## 1.2 数据集构建

目前中国没有可供使用的完整酒花数据集,所以通过自建平台采集白酒酒花图像构建数据集<sup>[19]</sup>。为了实现精确的酒花图像分类,采用控制变量法采集数据。使用环形打光,滤网(接酒面)放置于生产现场出酒口正下方20 cm处,酒液在接酒面中冲击出酒花,并通过滤网流入接酒桶,相机在滤网正上方垂直向下拍摄。现场出酒速度较快,酒花的形成和消散迅速,为了拍摄清晰的酒花图像,减少运动模糊,相机的快门速度要快。选用1 200万像素苹果高清相机,视频帧数为60帧/s,相机镜头位于接酒碗正上方50 cm处。滤网孔径大小直接影响气泡的成型,孔径过大则气泡无法在滤网上成型,孔径过小则会导致酒液下渗产生障碍,影响拍摄和生产。采集装置如图2所示。



1. 采集装置主体 2. 移动底座 3. 移动滚轮 4. 放置槽 5. 折叠三脚架 6. 开关盖 7. 接酒装置 8. 漏斗状滤网 9. 气孔 10. 酒篓 11. 第一竖杆 12. 弧形支撑 13. 流酒管 14. 出酒口 15. 智能电磁阀 16. 第二竖杆 17. 第一转动轴 18. 第一可调节转动限位杆 19. 相机夹 20. 摄像相机 21. 第二转动轴 22. 第二可调节转动限位杆 23. 环形补光灯

图2 图像采集平台示意图

Figure 2 Image acquisition platform

## 2 图像预处理

如图3所示,相机直接拍摄到的原图像中,有效的为接酒网部分区域,检测出这部分前景区域,舍弃背景部分,可以为后续步骤节约算力,加快速度。此外,生产现

场环境会有其他仪器引起的震动,使得相机拍摄可能产生偏移,前景部分也会发生偏移变化,因此需要实时检测。

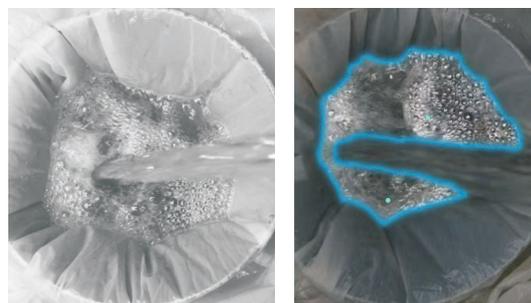


图3 自建平台采集的酒花气泡图像及需要的酒花区域  
Figure 3 Image of foams collected by and foam area required by self-built platform

由于试验使用自建平台,酒花采用400目滤网为接酒网,导致酒花图像边缘不规则,同时有水流部分形成的干扰,且酒花区域与滤网颜色较为接近,无法采用传统算法色彩空间HSV(Hue, Saturation, Value)进行滤波分离,因此使用基于语义分割的前景区域分割算法。为满足实时需求,采用轻量级语义分割网络ENet进行预处理。

### 2.1 ENet

ENet是一个针对语义分割任务设计的卷积神经网络模型,其核心目标是在保持高效的计算和低延迟的同时,提供高质量的分割结果,特别适用于移动和嵌入式设备。

ENet网络模型通过采用轻量级的卷积块和优化后的下采样方法,减少了参数数量,降低计算复杂度。ENet在设计时特别考虑了实时性能,使其在实时视频流中的应用变得可行,是能够达到实时分割的轻量网络,且只有37万的参数,与其原型算法SegNet的对比见表2。此外,ENet使用了深度可分离卷积等先进技术来进一步减轻模型负担,提高处理速度。

表2 SegNet与ENet参数量对比

Table 2 Comparison of parameter quantities between SegNet and ENet

模型	每秒浮点运算/ 亿次	参数量/M	模型大小(浮点 16位)/MB
SegNet	2 860.3	29.46	56.2
ENet	38.3	0.37	0.7

ENet通常包含3个主要部分:初始块(用于快速下采样)、一系列编码器块(用于提取并编码图像特征)和相应的解码器块(用于将编码的特征映射回原始图像分辨率以生成分割图)。这种架构使得ENet在多种场景下都能有效工作,包括道路场景理解和医学图像分割等应用<sup>[20]</sup>。

ENet 的网络结构见表 3。

表 3 ENet 网络结构图  
Table 3 ENet network structure

名称	操作类型	输出尺寸
Initial		16×256×256
Bottleneck1.0	downsampling	64×128×128
4*bottleneck.x		64×128×128
Bottleneck2.0	downsampling	128×64×64
Bottleneck2.1		128×64×64
Bottleneck2.2	dilated 2	128×64×64
Bottleneck2.3	asymmetric 5	128×64×64
Bottleneck2.4	dilated 4	128×64×64
Bottleneck2.5		128×64×64
Bottleneck2.6	dilated 8	128×64×64
Bottleneck2.7	asymmetric 5	128×64×64
Bottleneck2.8	dilated 16	128×64×64
Repeat section 2, without bottleneck2.0		
Bottleneck4.0	upsampling	64×128×128
Bottleneck4.1		64×128×128
Bottleneck4.2		64×128×128
Bottleneck5.0	upsampling	16×256×256
Bottleneck5.1		16×256×256
fullconv		C×512×512

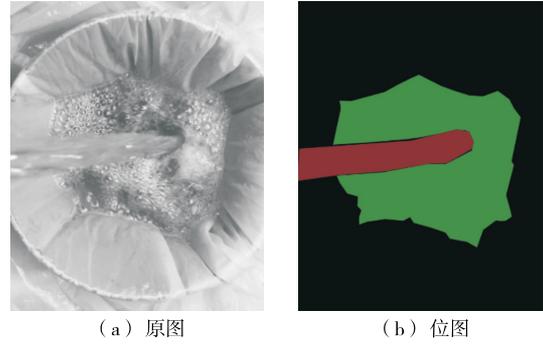
### 2.2 训练数据准备

试验实用场景中酒花气泡图像大部分阶段气泡非常密集且相互堆叠,气泡持续时间短,状态不稳定,导致很多气泡的边缘模糊,所以人工标注每一个气泡或每一堆气泡无疑会产生较大的误差。因此,使用目标分类方法,在人工标注阶段只需要将整体酒花区域轮廓标注即可,可以将人工误差降到最低。

首先用 labelme 标注软件标注图片,在每张图片中单独标出酒花区域和水流区域。每张图片生成对应的 .json 文件,并保存酒花和水流的坐标区域。通过运行 json2img.py,生成八位图(单通道的灰度图像)的 label,也称 mask,此 label 只有 3 个像素值即 0/1/2,0 代表背景,1 代表水流,2 代表酒花。神经网络通过生成的位图和 mask 进行对比,计算损失函数,通过梯度下降来更新网络参数,在不断地迭代中,神经网络将会学习如何分割水流和酒花的区域。图 4 为原图以及原图所对应的位图,位图显示为彩色是因为 png 图片中有一个 color 映射 cmap,所以显示为彩色,方便观测。

### 2.3 训练模型

试验网络不需要识别 background,所以将 background 类设置为 unlabelled 类,计算损失函数和 mIOU 时均不考



(a) 原图 (b) 位图

图 4 图像原图及位图

Figure 4 Original image and bitmap

虑 background 类别。损失函数为交叉熵损失函数 CrossEntropyLoss,初始学习率为 0.000 5,学习率应用台阶递减, factor 为 0.5,即每 50 个 epochs 之后,学习率变为原来的 1/2。使用 L2 正则化, factor 为 0.000 2,输入图片大小为 960 像素×540 像素,均为原图的 1/4。运行 200~250 个 epoch,网络即可完成学习,4060 笔记本显卡加速下耗时 1 h。学习过程输出中最优 mIOU 结果为 0.979 0。训练集和验证集的酒花 IOU 图如图 5 所示。

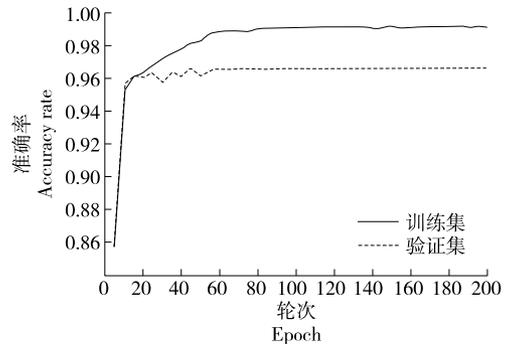


图 5 训练集和验证集的酒花 IOU 图

Figure 5 IOU charts of foams for training and validation sets

### 2.4 预测与输出结果

处理之后达到快速剥离前景部分的效果,在语义分割的实现过程中分割了酒花和水流两部分,水流部分一方面作为预处理部分的异常检测,如果没有水流检测为异常。另一方面水流部分形态与酒花不同会影响后续分类判断,所以进行检测之后就剥离掉,只保留酒花部分。预测及输出结果如图 6 所示。其中,960 像素×540 像素的图片输出速度可以达到 30 帧,满足实时需求。

## 3 酒花图像分类模型设计

在拍摄到的原始酒花图像中,真正有效的前景仅仅是接酒网所在区域,检测出该前景区域,并丢弃背景,可

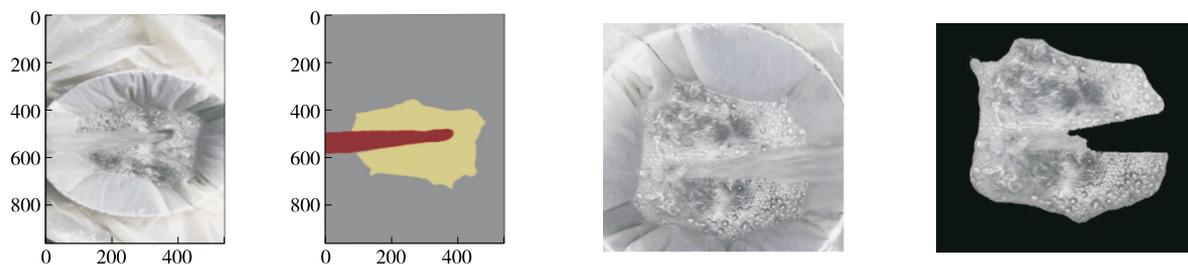


图6 图像预处理预测与输出结果图

Figure 6 Image preprocessing prediction and output results

以加快后续算法的运行。此外,摘酒过程中,相机可能会因为环境的震动而发生抖动或偏移,图像中接酒网的位置也会随之发生变化,因此需要对接酒网前景区域进行实时检测。

酒花分类算法以相机拍摄到的酒花图像作为输入,输出该图像对应的白酒的分段,并检测可能发生的异常,如图像遮挡、未流酒等。

在试验应用场景中,有的阶段气泡尺寸大如黄豆,有的阶段气泡小且极密集,有的阶段几乎没有成型气泡,需要从其他形态判断阶段,具体如图7所示。



图7 不同形态的酒花图像

Figure 7 Images of different foam morphologies

因此,试验不采用目标检测算法检测每一个气泡的位置和大小<sup>[21]</sup>,而是使用卷积神经网络算法提取包含有较多气泡的区域,对该区域图像进行整体的特征分类。

### 3.1 Vision Transformer

Vision Transformer (ViT) 应用了一种创新的方法,将传统的 Transformer 模型从文字处理成功迁移到图像处理领域(见图8)。它通过将图像分解成多个小块,每个块接着被转换成线性嵌入向量,这些向量按顺序排列,以此模仿自然语言处理中单词序列的输入方式。Vision Transformer的基本步骤包括图像的分割、平坦化处理、块嵌入、位置嵌入,最后通过 Transformer 编码器进行处理,并通过多层感知机(MLP)头部进行输出分类。

具体而言,ViT 从一个完整的图像中切割出等大小的小块。这些小块经过一个称为 Patch Embedding 的处理步骤,被转换成具有固定维度的向量。这一步骤涉及一次线性变换,其后还加入了一个额外的可学习的类别嵌入向量,用于提高模型对分类任务的处理能力。为了补充图像中位置信息的缺失,ViT 引入了位置编码技术,通

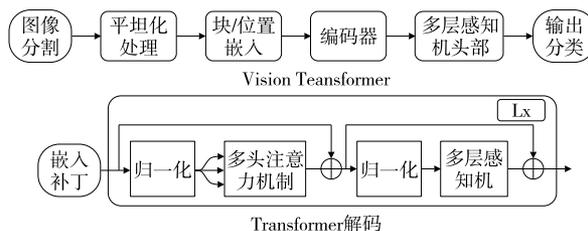


图8 Vision Transformer 整体结构示意图

Figure 8 Overall structure of Vision Transformer

过位置嵌入的方式增强了模型对物体在图像中位置的敏感性。Vision Transformer 基本流程为:图像分割→展平→块嵌入→位置嵌入→Transformer 编码器→多层感知机头部。

Vision Transformer 首先将一个  $H \times W \times C$  的图像变成一个  $x_p \in N \times (P^2 \times C)$  的展平二维块序列。为了得到 Transformer 希望的尺寸为  $(N, D)$  的输入矩阵,还需要对  $x_p \in N \times (P^2 \times C)$  的展平二维块序列进行 Patch Embedding 操作。Patch Embedding 操作将每个向量都做一次线性变换即可将展平二维块序列的维度压缩成  $D$ 。

$$z_0 = [x_{\text{class}}; x_p^1 E; x_p^2 E; \dots; x_p^N E] + Epos, \quad (2)$$

式中:

$Epos$ ——位置嵌入向量;

$E$ ——线性变换(即全连接层);

$P_2$  与  $C$ ——输入通道(二者相等);

$N$ ——输出通道;

$x_{\text{class}}$ ——额外增加的可学习的 class embedding。

后续的分类预测使用这个可学习的嵌入向量,而不使用  $N$  个展平二维嵌入向量,是为了避免需要选择哪个展平二维嵌入向量用于最后分类预测。

Vision Transformer (ViT) 通过引入位置编码(Positional Encoding)增强了模型对输入序列位置信息的感知能力。此外,ViT 特别设计了可学习的位置嵌入(Positional Embedding),用于在模型中显式地表示图像块之间的空间关系。在这种设计下,相互靠近的图像块通过其位置嵌入获得了相似性,这反映了图像的行列结构,使得同一行或同一列的图像块在特征空间中表现出较高

的相似度。完成位置嵌入后,模型进入 Transformer 编码器的核心处理(Transformer Encoder)阶段,核心处理过程为:

$$z'_l = \text{MSA}(\text{LN}(z_{l-1})) + z_{l-1}, l = 1, \dots, L, \quad (3)$$

$$z_l = \text{MLA}(\text{LN}(z'_l)) + z'_l, l = 1, \dots, L, \quad (4)$$

$$y = \text{LN}(z_L^0). \quad (5)$$

在 Transformer 编码器中,处理流程开始于多头自注意力机制(Multi-head Self-attention, MSA),该机制允许模型在处理输入序列时同时考虑不同子空间的信息。该步骤之后,模型执行多次的“加法与规范化”(Add and Norm)操作,这有助于维持训练过程中的数值稳定性并促进梯度传播。其次为前馈网络(Feed Forward Network),它进一步转换特征,并伴随更多的“加法与规范化”步骤。最终,经过一系列层处理后,首个 Patch 的输出通过多层感知机(MLP)和规范化步骤进行处理,从而产生最终的分​​类预测。

### 3.2 ConvNeXt

在传统的卷积神经网络基础上,ConvNeXt 结合了 Vision Transformer 中的一些新颖思想和技术,对经典的 ResNet 网络进行了一系列的调整和优化。这些优化措施包括但不限于采用新的训练技术、宏观设计的调整以及微观设计的优化。例如,ConvNeXt 在训练过程中采用了与 Vision Transformer 相同的训练技术,如使用 AdamW 优化器,这一策略关注于优化超参数的配置,以提升网络性能<sup>[22]</sup>。

在宏观设计方面,ConvNeXt 借鉴了 Swin Transformer 的设计理念,调整了 ResNet 中的层次次数配置,优化了计算资源的分配和利用。此外,ConvNeXt 采用了大卷积核和深层卷积结构,这些改动旨在通过更大范围的感受野捕捉图像特征,同时优化计算效率。在微观设计层面,ConvNeXt 通过采用 GELU 激活函数代替传统的 ReLU 激活函数,以及在网络正则化策略上使用 LayerNorm 代替 BatchNorm,从而提升了网络的稳定性和效率。其进行的具体优化设计主要有以下 6 点。

**3.2.1 Training Techniques (训练技术)** 在 Vision Transformer 中采用的创新训练技术不仅反映在其模块化设计和架构决策上,还包括多种新的训练方法,如使用 AdamW 优化器。这些方法关注于优化超参数的配置,从而显著提高模型性能。借鉴这一策略,ConvNeXt 采用了与 Vision Transformer 类似的训练流程来优化 ResNet50/200 的基准模型。

**3.2.2 Macro design (宏观设计)** ConvNeXt 在借鉴 Swin Transformer 的设计理念后,对 ResNet50 的层次结构进行了重组,将原有的层级分布(3,4,6,3)调整为(3,3,9,3)。在传统卷积网络中,初级下采样通常由 7×7 大小、步长为 2 的卷积层与步长为 2 的最大池化层组成,而

ConvNeXt 采取了与 Swin Transformer 相同的方法,使用一个大卷积核且不重叠窗口(如 4×4、步长为 4)的卷积层进行下采样,实现更加高效的特征提取。

**3.2.3 ResNeXt** 与经典的 ResNet 相比,ResNeXt 在运算量(FLOPs)和准确率之间找到了更好的平衡。ResNeXt 通过在卷积块中采用分组卷积,形成并行的结构设计,而传统的 ResNet 结构更倾向于瓶颈设计。ConvNeXt 进一步推进了这种设计,采用了更深层的卷积结构,以在略微牺牲准确度的前提下显著减少模型参数。

**3.2.4 Inverted bottleneck (倒置模块)** Transformer 网络中的 MLP 模块是“两头细,中间粗”的倒置模块。因此,ConvNeXt 网络也参照设计了一个类似的 Inverted bottleneck 结构(见图 9)。

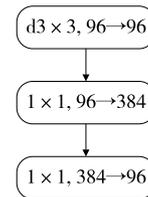


图 9 Inverted bottleneck 结构示意图  
Figure 9 Inverted bottleneck structure

**3.2.5 Large kernel size (大卷积核)** 在全局自注意力的基础上,ConvNeXt 调整了卷积核尺寸的策略,将深层卷积模块的配置从传统的“1×1 卷积—深度卷积—1×1 卷积”调整为先进行深度卷积,后接两次 1×1 卷积,这种结构调整模仿了 Transformer 中 MSA 模块前置的设计。此外,深度卷积的卷积核大小也从 3×3 调整至 7×7,以增强处理能力。

**3.2.6 Micro designs (微观设计)** 传统 CNN 中广泛使用的 ReLU 激活函数在 ConvNeXt 中被 GELU 激活函数替换,以配合 Transformer 架构的常见设置。此外,为了提高模型的整体效率和性能,ConvNeXt 在正则化技术上进行了调整,采用 LayerNorm 代替了常规的 BatchNorm,并简化了激活函数和正则化函数的使用。

ConvNeXt 的整体结构如图 10 所示。

## 4 试验设计及结果分析

### 4.1 酒花分类训练数据准备

试验共采集 12 条酒花视频数据,其中大青花、小青花、云花、水花各 3 条。相应的视频后每间隔 6 帧进行一次数据采集,以此来得到数据集中的图像。通过此方法,剔除存在图像遮挡、图像偏移、标签争议的图像后,共得到 9 746 张图像,其中训练图像 8 592 张,测试图像 1 154 张。

为了更好地展现通过 ENet 对数据的处理对分类准确

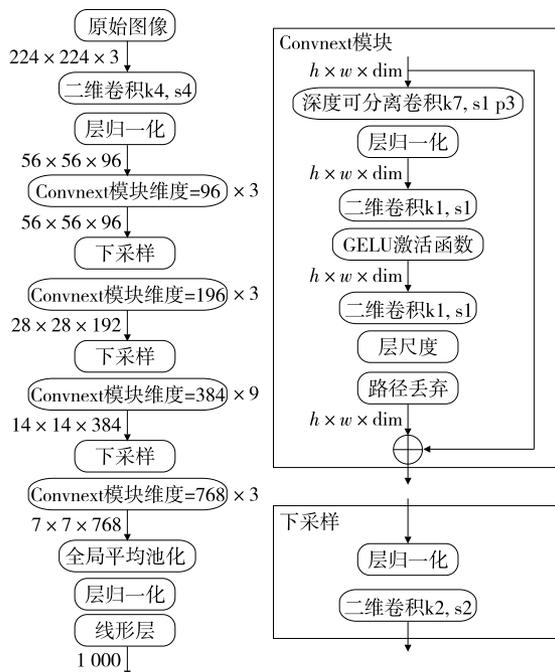


图 10 ConvNeXt 的整体结构示意图

Figure 10 Overall structure of ConvNeXt

率的提升效果,以相同方法,获取了原始视频的对应数据集,同样为 9 746 张图像,其中训练图像 8 592 张,测试图像 1 154 张。

#### 4.2 测试指标及测试方法

通过计算分类准确率来表示算法的整体精度。假设共有  $N$  个待测试数据,其中第  $i$  个数据的真实标签与预测标签分别为  $y_1, y_i^*$ ,则分类准确率按式(6)进行计算。

$$A_{cc} = \frac{1}{N} \sum_{i=1}^N I(y_i^* = y_i), \quad (6)$$

式中:

$A_{cc}$ ——准确率, %;

$I(x)$ ——指示函数,  $x$  为真时取 1,  $x$  为假时取 0。

除准确率外,还采用执行时间来描述算法的速度,即完成一张图像的分类所需的时间。为保证公平性,所有试验均在 GeForce RTX 2080Ti 11 G 服务器上进行,并将 batch size 设置为 4 用于训练。

#### 4.3 试验结果及相关分析

为了取得更好的训练结果,针对原始的训练数据进行数据增强,并使用进行数据增强后的数据对 Vision Transformer 与 ConvNeXt 进行相应的网络训练。数据增强主要采用 pytorch 自带的 RandomHorizontalFlip 方法,以 0.5 的默认概率水平翻转给定的图像。

为保证公平,所有的训练及测试数据均设置为  $224 \times 224$ 。由表 4 可知,ConvNeXt 对酒花的分类准确度达 98.4%, 优于 Vision Transformer 的,这归因于其基于

Vision Transformer 训练过程以及多方位的改进措施。而 ENet 的引入又使得 Vision Transformer 与 ConvNeXt 的分类准确率分别提高了 1.1%, 1.0%。

表 4 Vision Transformer 与 ConvNeXt 分类试验对比结果

Table 4 Comparison results of Vision Transformer and ConvNeXt classification experiments

分类模型	分割前准确率/%	分割后准确率/%
Vision Transformer	97.8	98.9
ConvNeXt	98.4	99.4

为增加说服力,通过可视化分类结果图的方式来展现 Vision Transformer 与 ConvNeXt 的性能差异以及 ENet 分割数据集对分类准确率的提升效果。由图 11 可知,ENet 很好地将水流以及周遭背景部分剔除,只留下较为关键的酒花区域部分。



(a) 分割前 (b) 分割后

图 11 ENet 分割前后对比图

Figure 11 Comparison before and after ENet segmentation

由图 12 可知,未分割前,VIT 错误地将酒花为大花的图片划分为小花,其原因在于小花与大花的关键区别在于泡沫大小,而此张图像边缘部分泡沫较小,且存在水流遮挡的干扰,因此产生了误判。分割后,ENet 去除了水流以及滤网部分的干扰,余下部分为较为明显的大花特征,因此正确分类。

由图 13 可知,虽然小花仍存在一定的置信度,但远低于大花的置信度,说明 ConvNeXt 相对于 VIT 性能更优越。

由表 5 可知,ConvNeXt 和 Vision Transformer 推理一张  $224$  像素  $\times$   $224$  像素的图片速度为 5.06, 7.07 ms, 均快于 EfficientNetf 的。对于 ConvNeXt 来说,其采用了交叉通道连接的方法来增强特征提取能力,这种设计会在推理阶段减少模型计算量,从而提高推理速度。而对于 Vision Transformer 来说,Transformer 的结构支持并行计算, GPU 对这种结构的加速效果更好。反观 EfficientNetV2,其网络结构中使用了大量的低 FLOPs、高

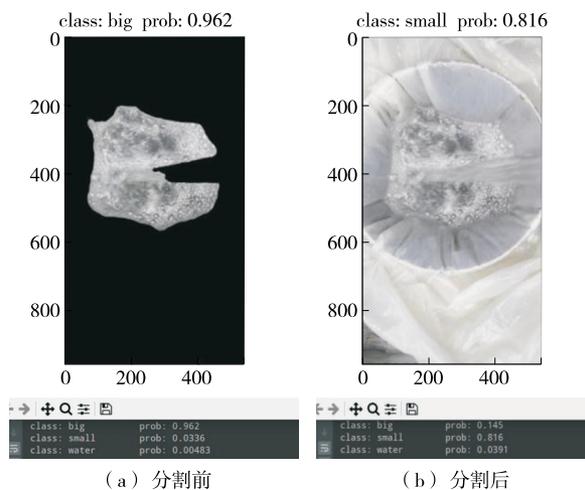


图 12 VIT 下 ENet 分割前后分类准确率对比图

Figure 12 Comparison of classification accuracy before and after ENet segmentation under VIT

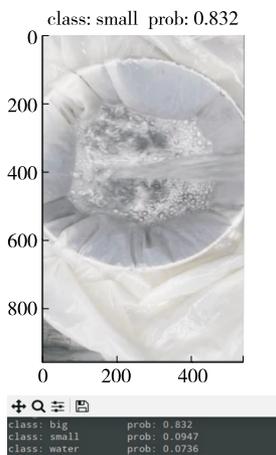


图 13 原图像在 ConvNeXt 下的分类准确率结果图

Figure 13 Classification accuracy results of original image under ConvNeXt

数据读写量的操作,即 depthwise 卷积操作,这些具有高数据读写量的操作,加上 GPU 的访存带宽限制,使得模型将大量的时间浪费在显存中读写数据上, GPU 的算力未得到充分利用。

## 5 结论

针对现有计算机视觉摘酒方案的实时性、准确性与

表 5 各模型运行时间汇总

Table 5 Summary of running time for each model

模型	运行时间/ms
EfficientNetV2 <sup>[23]</sup>	25.02
Vision Transformer	7.07
ConvNeXt	5.06

普适性不平衡的问题,提出了一种基于语义分割网络 ENet 和卷积神经网络 ConvNeXt 的白酒酒花实时检测方法。通过语义分割网络 ENet 取代传统算法的图像预处理,在满足实时的前提下,能适用于相对复杂且有波动的生产场景;并在酒花图像分类部分对于最新的 VIT 模型和 ConvNeXt 模型进行了对比分析,最终选取了卷积神经网络 ConvNeXt 作为分类算法模型。结果表明,试验提出的检测方法准确率达 99.4%,相比相关领域模型准确率有所提升;整体速度能达到 30 帧/s 的输出速度,足以满足实时需求,与现有酒花分类算法相比有较大优势。后续可以增加不同场景的数据集,以提升模型对较复杂场景的普适性,同时由于不同阶段酒花气泡持续时间即“坐花时间”也有明显区别,具体到每一帧图像中就是破裂气泡的占比不同,因此同样可以成为改进模型的研究方向。

## 参考文献

- [1] 毛洪川. 整粒高粱酿造浓香型白酒的工艺开发及应用研究 [D]. 绵阳: 西南科技大学, 2017: 117-122.  
MAO H C. Development and application of a new technology of Luzhou flavour Liquor using intact sorghum as raw material [D]. Mianyang: Southwest University of Science and Technology, 2017: 117-122.
- [2] 马娜. 荧光阵列传感器的构建及在白酒风味物质检测中的应用研究 [D]. 重庆: 重庆大学, 2018: 16-21.  
MA N. Fabrication of fluorescent system and its application in the detection of liquor flavoring substances [D]. Chongqing: Chongqing University, 2018: 16-21.
- [3] 刘智萍, 崔克彬. 基于深度学习的白酒酒花实时分类方法 [J]. 食品与机械, 2022, 38(11): 111-116.  
LIU Z P, CUI K B. Real-time classification method for liquor hops based on deep learning [J]. Food & Machinery, 2022, 38 (11): 111-116.
- [4] 泸州老窖集团有限责任公司. 泸州酒技艺大全 [M]. 北京: 中国轻工业出版社, 2011: 105-106.  
Luzhou Laojiao Group Co Ltd. Complete collection of Luzhou type liquor techniques [M]. Beijing: China Light Industry Press, 2011: 105-106.
- [5] 何超, 余东, 李战国, 等. 近红外光谱技术在量质摘酒中的应用探索 [J]. 酿酒, 2019, 46(4): 58-60.  
HE C, YU D, LI Z G, et al. Application of near infrared spectroscopy to gathering distillate according to the quality [J]. Liquor Making, 2019, 46(4): 58-60.
- [6] WU Z Z, XU E B, LONG J, et al. Measurement of fermentation parameters of Chinese rice wine using Raman spectroscopy combined with linear and non-linear regression methods [J]. Food Control, 2015, 56: 95-102.
- [7] 杨静娟, 任小洪. 基于多纹理特征的白酒摘酒酒花图像分类

- 识别[J]. 包装与食品机械, 2021, 39(4): 38-45.
- YANG J X, REN X H. Classification and recognition of liquor receiving hops images based on multiple texture feature extraction[J]. Packaging and Food Machinery, 2021, 39(4): 38-45.
- [8] 潘斌, 韩强, 姚娅川. 基于卷积神经网络的白酒酒花分类研究[J]. 食品与机械, 2021, 37(10): 30-37, 88.
- PAN B, HAN Q, YAO Y C. Research on Classification of liquor hops based on convolution neural network[J]. Food & Machinery, 2021, 37(10): 30-37, 88.
- [9] 余锴鑫. 基于图像分类算法的自动化摘酒方法研究[D]. 杭州: 浙江大学, 2019: 10-12.
- YU K X. Research on automatic wine picking method based on image classification algorithm[D]. Hangzhou: Zhejiang University, 2019: 10-12.
- [10] 田子宸. 基于计算机视觉的智能摘酒方法研究[D]. 杭州: 浙江大学, 2020: 60-67.
- TIAN Z C. Research on intelligent liquor selection method based on computer vision[D]. Hangzhou: Zhejiang University, 2020: 60-67.
- [11] 王继华, 李兆飞, 杨壮, 等. 面向不平衡数据集的浓香型白酒基酒等级分类研究[J]. 中国酿造, 2024, 43(1): 184-189.
- WANG J H, LI Z F, YANG Z, et al. Research on grade classification of strong-flavor Baijiu base liquor based on unbalanced data sets[J]. China Brewing, 2024, 43(1): 184-189.
- [12] 朱琳, 刘恩, 蒋文, 等. 乙醇水溶液表面张力的模型拟合[J]. 数理医药学杂志, 2008, 21(6): 647-649.
- ZHU L, LIU E, JIANG W, et al. Model fitting of surface tension of ethanol aqueous solution[J]. Journal of Mathematical Medicine, 2008, 21(6): 647-649.
- [13] 刘江生, 贾怡娴, 韩青梅, 等. 酒花与酒的探讨[J]. 酿酒, 2014, 41(3): 36-41.
- LIU J S, JIA Y X, HAN Q M, et al. Discussion on the relationship between liquor whirlpool and liquor[J]. Liquor Making, 2014, 41(3): 36-41.
- [14] 张琴, 银莹, 杨宛宜, 等. 清香型白酒中甲醇、乙酸乙酯和乳酸乙酯的检测[J]. 中国酿造, 2023, 42(11): 262-266.
- ZHANG Q, YIN Y, YANG W Y, et al. Detection of methanol, ethyl acetate and ethyl lactate in light-flavor Baijiu[J]. China Brewing, 2023, 42(11): 262-266.
- [15] 方昕. 聚乳酸的乳酸乙酯分散液制备及其在纺织上的应用[D]. 上海: 东华大学, 2014: 25-33.
- FANG X. Preparation of polylactic acid dispersion with ethyl lactate and its application in textiles[D]. Shanghai: Donghua University, 2014: 25-33.
- [16] 康乐, 冯殿义, 胡榕华. 液体泡沫中单气泡的图像识别算法[J]. 渤海大学学报(自然科学版), 2017, 38(2): 167-172.
- KANG L, FENG D Y, HU R H. Single bubbles recognition algorithm in liquid foam image[J]. Journal of Bohai University (Natural Science Edition), 2017, 38(2): 167-172.
- [17] 张琦, 沈才洪, 孙啸涛, 等. 浓香型白酒蒸馏过程中挥发性风味物质的变化规律[J]. 现代食品科技, 2018, 34(6): 244-254.
- ZHANG Q, SHEN C H, SUN X T, et al. Variation regularity of volatile substances in Luzhou flavor liquor during distillation process[J]. Modern Food Science and Technology, 2018, 34(6): 244-254.
- [18] 熊子书. 试论白酒香味成分与风味的关系[J]. 黑龙江发酵, 1980(3): 10-13.
- XIONG Z S. On the relationship between flavor components and flavor of Baijiu[J]. Heilongjiang Fermentation, 1980(3): 10-13.
- [19] HOLLOWAY D C, JPHNSON R H. Advancements in holographic photoelasticity[J]. Experimental Mechanics, 1971, 11(2): 57-63.
- [20] PASZKE A, CHAURASIA A, KIM S, et al. ENet: a deep neural network architecture for real-time semantic segmentation[J]. Computer Science, 2016, 14(12): 77-79.
- [21] 崔涛. 浅谈白酒中的高级醇[J]. 酿酒科技, 1994(6): 28-29.
- CUI T. Elementary introduction to higher alcohols in Baijiu[J]. Liquor-Making scence & Technology, 1994(6): 28-29.
- [22] LIU Z, MAO H, WU C Y, et al. A ConvNet for the 2020s[C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA USA. IEEE, 2022: 77-82.
- [23] MA N, ZHANG X, ZHENG H T, et al. ShuffleNet V2: practical guidelines for efficient CNN architecture design[C]// Computer Vision-ECCV 2018. [S. l.]: Springer International Publishing, 2018: 122-138.