

# 增值税发票图像处理与文本定位技术研究\*

张浩晨 李陶深\*\*

广西大学计算机与电子信息学院, 南宁, 530004

**摘要** 对增值税发票处理与文本定位技术进行研究。在严格保护数据隐私的情况下, 对所收集的纸质增值税发票图片进行预处理主要包括图像倾斜矫正、灰度化、将 RGB 图像转为 HSV 空间, 并对其中的噪音、干扰信息等进行消除, 对增值税发票中的重要信息保留并增强处理。基于连通域的文本定位方法, 对预处理后的图像进行重要信息文本的具体定位, 并对在预处理阶段未被处理的干扰信息进行二次消除, 从而得到需要进行文本识别的区域信息, 为后续文本识别做准备。

**关键字** 水果图像识别, 深度学习, 卷积神经网络, MobileNet V2 (depthwise separable convolution)

## Research on Image Processing and Text Location Technology of Value Added tax Invoice

Zhang Haochen Li Taoshen

School of Computer, Electronics & Information  
Guangxi University  
Nanning 530004, China  
tshli@gxu.edu.cn

**Abstract**—The value-added tax(VAT) invoice pre-processing and text location technology are studied. Firstly, under the condition of strictly protecting data privacy, preprocessing the collected paper VAT invoice images mainly includes image tilt correction, grayscale, converting RGB images into HSV space, and noise and interference information. etc. which can to eliminate, retain and enhance processing of important information in VAT invoices. After preprocessing, image information that is easy to locate text is obtained. Secondly, the text localization method based on the connected domain is used to locate the important information text of the preprocessed image, and the interference information that has not been processed in the preprocessing stage is eliminated twice, so as to obtain the region information that needs to be recognized by the text.

**Key words**—Fruit image recognition, convolutional neural network, deep learning, MobileNet V2

## 1 引言

经济社会的高速发展和生活水平的提高, 人们对水果的种类及品质的要求越来越高。但是水果种类繁多, 因此水果的分类检测就显得十分必要。传统的水果分类方法主要是用人工方法, 费工、费时, 人力成本高, 且成本效益较低。一些研究人员在水果识别和分类领域开展了技术攻关和应用合作, 取得了较好的技术研究成果<sup>[1]</sup>。在实际应用中, 人们开发了不同的水果识别系统, 积累了大量的水果图像数据资源。

当前, 对水果分类方式的研究大多采用了傅里叶变换和水果表面颜色。前者先测量彩色水果图形的边界, 进而获取水果的轮廓 0.5 零点五径范围序列并进

行离散傅里叶转换。最后再利用傅里叶系数定义分类器, 并按照所给出的划分阈值对水果的外形特征进行类型划分。后者通过将苹果图片转换为 HIS 颜色模型后, 对苹果图片中 H 分量的面积加权直方图的二个主要成分加以分析, 从而进行了水果分类, 分级误差仅为 1.75% 百分之一七五。2008 年李甦等人<sup>[2]</sup>总结了生物目标表面颜色的分布变化规律, 并分别构建出“数量—程度”和“尺寸—评价光度值”的空间模式。前者能精确说明果面的程度, 后者能精确地说明水果种类。

最近几年, 深度学习在果蔬采摘和水果分类方面得到了较多的应用<sup>[3]</sup>。在水果采摘方面, 伍锡如等人<sup>[3]</sup>应用深度学习方法, 设计了一种水果采摘机器人视觉识别系统, 可以提高水果识别的准确性和实时性。牟其松<sup>[4]</sup>以苹果、桃子、香橙、梨为研究对象, 将深度学习与定位技术结合起来, 提出了基于深度学习的水果采摘通用检测模型。杨平<sup>[5]</sup>提出了一种改进的水果

\*基金资助: 本文得到广西高等教育本科教学改革工程项目一般项目(2020JGA116)资助。

\*\*通讯作者: 李陶深, 男, 教授, tshli@gxu.edu.cn

采摘目标识别算法,解决了小果实和被遮挡的果实无法识别的问题。在水果分类方法的研究方面, Hetal 等人<sup>[1]</sup>提出了基于多特征的检测水果对象算法。尼纳维等人<sup>[1]</sup>描述了一种水果分类方法,该方法使用具有图像纹理的 K-最近邻(KNN)分类器作为新特征,提高了水果分类的效率。李武提出了一种基于多类支持向量机的水果分类方法形式。何江萍等人<sup>[6]</sup>提出了一种基于视觉特征的水果蔬菜自动分类方法,实现了超市中水果蔬菜等产品的自动销售。李冲冲<sup>[7]</sup>以皇冠梨、橙子和苹果为研究对象,设计了一套基于机器视觉的水果检测系统,研究和改进了基于机器视觉的水果表面缺陷检测算法,提高水果检测效率与准确率。许学斌等人<sup>[8]</sup>提出了一种基于 LBP 的粒子群优化混合核函数 ELM 的蔬菜水果图像分类方法,提高了水果分类的准确率,可以满足智慧农业的实际需求。

本文引入深度学习方法,研究基于深度学习的水果种类识别技术方案,以便提高水果分类的准确率,满足智慧农业的实际需求。本文基于深度学习的水果种类识别技术方案的实现步骤大致分为数据预处理、模型的构架与训练、水果图像识别等三个部分。

## 2 数据预处理

### 2.1 数据集的准备

数据是深度学习研究的基础,图像获取是图像识别的第一步,得到图像后要进行处理。高清、高质量的图像样本和样本数量对于训练一个好的深度学习模型是非常重要的。

经过仔仔细细的挑选,我们最终选取到梨子、芒果、苹果、圣女果、香蕉等 5 种水果作为本文技术研究的水果图像数据集。

### 2.2 数据处理

常用的数据处理方法有:零均值化、数据归一化、主成分分析法等。本文采用主成分分析法。

主成分分析(Principal Component Analysis, PCA)是一种无监督学习方法,也是一种最常用的数据降维方法。它的主要任务是在原始空间中随机找到一个相互正交的坐标轴,而新的旋转坐标轴的选择也与原始数据本身有关。其中第一个新轴被选为原始数据上方变化最大的位置,而第二个新轴被选为与第一个轴正交的平面中变化最大的位置,第三个轴是和第一个轴。方差在两个轴正交的平面中最大。以此类推,可以得到  $n$  个这样的新坐标轴。对于这样得到的新坐标轴,可以看出前  $k$  个坐标轴的方差大部分都包含在内,而后面的坐标轴的方差基本为零。这样,可以省略其余的轴,只留下方差最大的前  $k$  个轴。但实际上,这意味着只留下方差最大的维度特征,而将方差基本为零的特征

维度略去,以便于对数据特征进行降维分析。

### 2.3 数据增强

数据增强是扩展大数据样本规模的有效方式。人们目前期望规模越大,数据品质就越高越好。这种模式可能具备良好的泛化能力。但是,在实际收集数据的时候,往往无法涵盖任何场景。例如针对光照条件下,在收集图像数据时很难控制光照的比例。因此,在训练建模时,需要在光照变化方面加入数据增强。另一方面,数据的获取也需要大量的成本。如果能够自动生成各种训练数据,将有可能实现更好的开源和高性价比<sup>[10]</sup>。

图1是水果图片数据处理经过图像增强后的图片对比图。图的左上方小图是原图,而右下方的小图则是对原图片进行随机剪切、压缩、翻转等运算后所得的。处理时,为了提高图像的质量,我们数据扩展了 10 倍。假设一个输入到网络上的图象的分辨率大小为  $256 \times 256$ ,如果随意裁剪大小为  $224 \times 224$ ,则一个图象最多能够得到  $32 \times 32$  位的图象,数据量就被放大了近 1000 倍。但是因为很多图的相似度都比比较高,与实际效果并不等价。因此,如果补充其他的数据增强方案,会获得更多的数据集,这正是数据增强的实质。

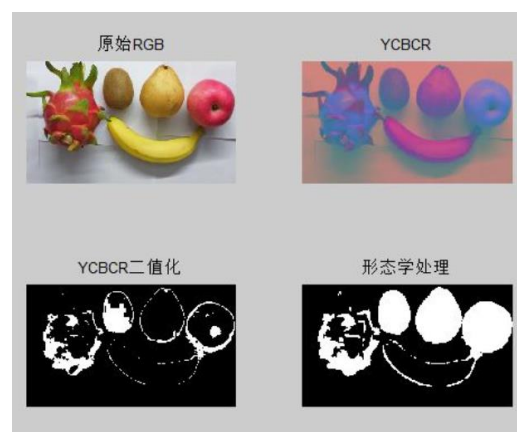


图 1 水果图片数据处理图片对比图

一般来说,构建同时产生简单和快速算法的数据增强方案是一门艺术问题,因为成功的策略与所考虑的(观察数据)模型有很大差异。我们采用了文献[11]介绍的一种有效的搜索策略,该策略结合了边缘增强和条件增强的思想,以及用于选择良好增强方案的确定性近似方法。

### 2.4 图像镜像处理

在深度学习中,图像镜像是指跨垂直或水平轴创建图像的反向副本的过程。水平翻转图像将产生镜像反射效果,而垂直翻转图像将类似于物体在水中的反射,也称为水反射效果。本文以图片对称轴为中心对图片进行镜像处理,即以图象的横轴与纵轴的交点为中心,

对图像进行镜像。图2给出了不同镜像处理的效果图。

经过上述数据增强方法处理后,本文使用的数据集就创建好了。该数据集包含5类水果图像:梨、芒果、苹果、樱桃番茄和香蕉,共311张图片。图2则是水果图像数据集数据的部分样本展示,其中第一行是梨,第二行是芒果,第三行是苹果,第四排是圣女果,第五排是香蕉。

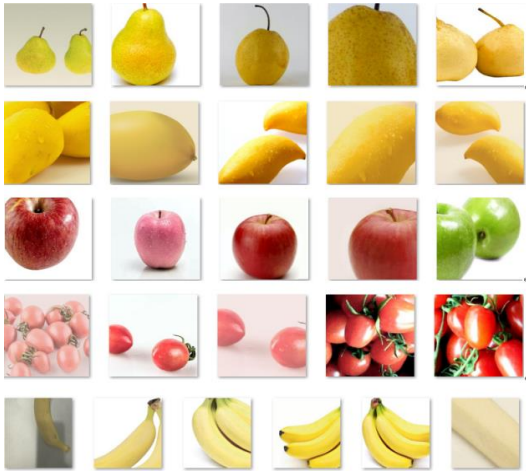


图 2 图像数据集的部分展示

### 3 基于 CNN 的水果图像识别技术研究

#### 3.1 CNN 简介

卷积神经网络(Convolutional Neural Network, CNN)是多层感知器的一个变种。它是根据美国生理学家 Huber 和 Wiesel 对动物视觉皮质的早期研究成果发展起来的。动物视觉皮质的神经细胞拥有着复杂的构造,这种细胞对于视觉输入位置的原子区特别灵敏,人们称为感知野细胞,它以这种形式平铺了整个视觉。感知野细胞又可分成二个基本形式,单纯细胞和复杂细胞。单纯细胞对于感知野内的边缘刺激模式作出最大反馈。而复杂细胞则产生了最大的感知野,这种感知野对来自确切区域的刺激是局部不变的<sup>[12]</sup>。

CNN 也是一个前馈神经网络,可以在有卷积结构的数据中获得特征。CNN 的架构中,一个生物神经元可以对应一个人工神经元,内核则代表着能够响应不同特性的不同受体;激活函数模拟了只有在达到某个阈值时的神经电信号才传导给下一神经元的功能。而损失函数和优化器则是由人类所独创的,目的是整 CNN 系统学习我们期望的东西<sup>[13]</sup>。目前, CNN 已经被广泛应用于分类识别多维度的图像中的平行移动、伸缩放大和其他形式的高度失真不变性。在该网络中特征提取参数是从训练数据中汲取出来的,在某种程度上规避了手动提取特征的必要。

#### 3.2 轻量级神经网络 MobileNet V2

MobileNets 是谷歌公司提供的一款用于手机等嵌入式电子设备上的轻量级深度神经网络,它在结构上采用了与非流线型相对应的流线型整体结构,使得用户在考虑复杂度时能够更合理地均衡延迟和精确度。

MobileNets 网络模型使用 Inception 模型来降低减少前几层的庞大的计算量。平扁化的网络结构使得我们能从完全分离的卷积中构建一个轻量级网络,演示了极度分离网络的潜力 **Error! Reference source not found.** MobileNet 模型可以应用于各种识别任务,以提高设备智能的效率。

MobileNet V2 是谷歌继 MobileNet V1 之后推出的下一代轻量级网络,克服了 V1 在训练过程中容易出现特征退化的问题。MobileNet v2 主要特性有三个:一是使用深度可分离卷积代替普通卷积,以降低模型的计算量和参数量;二是提出反向残差结构加深网络层数和增强特征的表达能力;三是采用线性瓶颈结构代替非线性瓶颈,减少低维特征信息的丢失。

MobileNet V2 的结构正好与传统 residual block 中降维后扩展相反,所以 shortcut 连接降维后就变成了 feature map。MobileNet V2 的核心模块主要由 Inverted residuals 与 Linear Bottlenecks 组成。

##### (1) Inverted residuals

残差结构如图 3 所示。这种反向中残差结构首先经过  $1 \times 1$  卷积运算将通道数压缩,然后再通过  $3 \times 3$  卷积运算,最后再通过  $1 \times 1$  卷积运算实现扩张;而反向残差网络 channel 为先扩张后压缩。引入 shortcut 是为提高梯度在多层之间的传播能力,防止梯度消失。这种反向残差结构可以提高内存利用率。

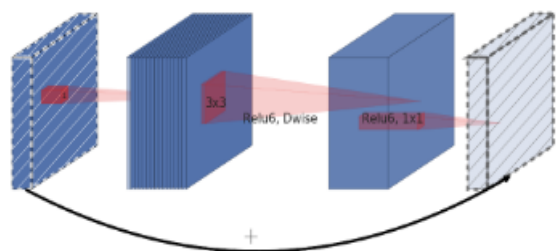


图 3 反向残差结构示意图

##### (2) Linear Bottlenecks

经过反向残差模块,输出的 feature map 经过 channel 压缩,损失一部分特征,如果使用 relu 激活函数,会损失更多特征,因此采用线性变换,即 Linear Bottlenecks。图 4 给出了 Bottleneck 单元结构示意图。



Input	Operator	$t$	$c$	$n$	$s$
$224^2 \times 3$	conv2d	-	32	1	2
$112^2 \times 32$	bottleneck	1	16	1	1
$112^2 \times 16$	bottleneck	6	24	2	2
$56^2 \times 24$	bottleneck	6	32	3	2
$28^2 \times 32$	bottleneck	6	64	4	2
$28^2 \times 64$	bottleneck	6	96	3	1
$14^2 \times 96$	bottleneck	6	160	3	2
$7^2 \times 160$	bottleneck	6	320	1	1
$7^2 \times 320$	conv2d 1x1	-	1280	1	1
$7^2 \times 1280$	avgpool 7x7	-	-	1	-
$1 \times 1 \times k$	conv2d 1x1	-	k	-	-

图 7 基于MobileNet V2的水果分类与识别的网络架构

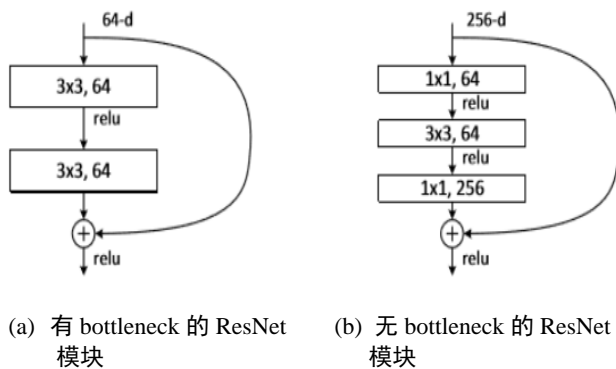


图 8 有或没有 bottleneck 的 ResNet 模块

## (2) 模型的训练

模型训练的主要步骤如下:

① 数据集加载函数,指明数据集的位置并统一处

理为  $imgheight \times imgwidth$  的大小,同时设置 batch:  

```
def data_load(data_dir, test_data_dir, img_height, img_width, batch_size)
```

② 构建mobilenet模型:训练模型并进行数据加载,进行图片处理,即图片的大小及进行迁移学习。

```
def model_load(IMG_SHAPE=(224,224,3), class_num=12)
```

微调的过程中是不需要考虑进行归一化的处理。

③ 展示训练过程的曲线:

```
def show_loss_acc(history)
```

从history中提取模型训练集和验证集准确率信息和误差信息:

④ 开始训练,记录开始时间

```
def train(epochs):
```

```
begin_time = time()
```

## (3) 模型的优化

多数机器学习算法都是建立一个算法训练模型,并使用优化方法对目标函数进行优化,以训练出理想的模型。常用的优化方案有梯度下降法、牛顿法、拟牛顿法、共轭梯度法等。我们采用的是梯度下降法和

共轭梯度法。

## 4 结束语

本文研究了水果分类与识别算法的相关技术,提出了一种基于深度学习的水果种类识别技术方案。该方案首先是对水果的图像数据进行预处理,建立了5类水果的数据集。其次,对水果图像进行数据增强和镜像处理,以便获得质量较高的图像样本及样本数量,便于下一步对深度学习模型的训练。最后,利用轻量级卷积神经网络 MobileNet V2 构建基于卷积神经网络化的水果图像识别模型,并通过实验和训练,发现和挖掘大量数据中的空间关系,在一定程度上降低网络中的训练模型的参数含量,进而提高反向传播算法的有效性。利用上述的技术方案和我们构建的水果数据集,我们实现了一个基于深度学习的水果种类识别的样机系统。系统的识别测试结果表明,系统对五类水果的分类和识别具有接近96%的准确率,说明了本文提出的技术方案的可行性。

## 参考文献

- [1] 许伟,周星星,张煜东.水果分类方法研究综述[J].电子技术与软件工程,2015,(17):76-77
- [2] 李甦,谭永龙,杨美英.水果分级与表面缺陷检测研究[J].计算机工程与设计,2008,(15):3954-3957
- [3] 伍锡如,雪刚刚,刘英璇.基于深度学习的水果采摘机器人视觉识别系统设计[J].农机化研究,2019,42(2):177-182,188
- [4] 牟其松.基于深度学习的水果采摘通用检测模型研究[D].天津:天津理工大学,2021.03
- [5] 杨平.基于深度学习的水果采摘定位算法的研究[J].哈尔滨:哈尔滨理工大学,2020
- [6] 何江萍,马彦,李强.基于视觉特征的水果蔬菜自动分类方法[J].重庆师范大学学报(自然科学版).2016,33(03):115-120
- [7] 李冲冲.基于机器视觉的水果分类方法研究[M].郑州:河南工业大学,2019.
- [8] 许学斌,赵雨晴,路龙宾等.基于LBP的粒子群优化混合核函数ELM的蔬菜水果图像分类方法[J].机械设计与研究.2021,37(04):15-20,25.
- [9] Hinton G E,Osindero S,Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J] Neural Computation,2014,18(7):1527-1554.
- [10] Bengio Y,Delakkeau O. On the expressive power of deep architectures[C]. Algorithmic Learning Theory,Berlin Heidelberg,2011:18-36.
- [11] 张慧,王坤峰,王飞跃.深度学习在目标视觉检测中的应用进展与展望[J].自动化学报,2017,43(8):1289-1305.
- [12] Hinton G E,Osindero S,Teh Y W . A Fast Learning Algorithm for Deep Belief Nets[J] Neural Computation,2014,18(7):1527-1554.
- [13] He K,Zhang X,Ren S,et al.Deep Residual Learning for Image Recognition[J].2015.
- [14] Dong T Y.A Simple Analysis of AlphaGo[J]. Acta Automatica Sinica,2016.