

基于深度神经网络的违禁食品检测系统设计与实现

孙鑫宇 杜晓龙 陈宇林 邵世智

(南京科技职业学院信息工程学院, 江苏 南京 211000)

摘要: 校园重点场所如教学楼、机房和实验室等严禁带入违禁物品是众所周知的规则, 因此, 违禁物品的识别与分析处理是智慧校园治理亟待解决的问题。本文以违禁食品进教学楼的应用场景切入, 基于深度神经网络的人工智能处理技术设计实现了智能检测系统, 搭建了实验平台, 利用 TensorFlow 内置 Keras 模块构建卷积神经网络, 实现对目标范围内人、食品进行图像抓取, 智能识别图像内容, 分析图像中人的身份、食品的种类并进行智能决策, 根据决策结果进行自然语言合成播报提醒信息, 并将识别和决策结果上传校园云信息存储中心, 提升校园的智能化治理水平。

关键词: 违禁食品识别; 深度学习; 卷积神经网络; 自然语言合成

违禁物品识别与曝光处理是常见的应用领域之一, 诸如车站、机场违禁物品检查、动植物的拍照识别、门脸确认和垃圾智能识别与分拣等都属于该领域。目前针对这类违禁品的检查常用的是基于 X 射线扫描技术实现对安检图像中特定目标的识别, 然而这种检查方法不仅需要大量硬件设备投入, 而且依赖人力配合检查扫描结果, 效率低。因此, 如何利用并优化目前先进的人工智能领域的图片或视频智能识别技术, 实现设备自动、无接触检测, 并实现检测结果的可保存、可追溯和可集成应用是亟待解决的一个问题。

本文以校园违禁食品进楼宇的识别与分析处理为切入场景, 基于深度神经网络的人工智能处理技术设计实现了智能检测和监督系统。深度学习具有学习能力强、覆盖范围广、适应能力强、可移植性好等特点, 使其在图片识别的领域中得到了广泛应用, 随着深度学习技术在目标检测和图像处理领域的深入研究, 卷积神经网络 (Convolutional Neural Network, CNN) 在图像识别和分类中的应用日益广泛, 国内外开展了大量针对卷积神经网络在目标识别领域的研究工作, 尤其在食物识别研究中, 卷积神经网络被证明比传统机器学习方法具有更好的性能。本系统搭建了实验平台, 利用 TensorFlow 内置 Keras 模块构建卷积神经网络的, 实现对目标范围内人、食品进行图像抓取, 智能识别图像内容, 分析图像中人的身份、食品的种类并进行智能决策, 根据决策结果进行自然语言合成播报提醒信息, 并将识别和决策结果上传校园云信息存储中心, 实现校园的智能化治理水平。

一、系统设计

本系统基于深度学习的多级卷积神经网络算法设计了集成框架模型, 采用无接触式的拍照检测, 完成目标图像的视觉识别、分析和自然语言提醒信息合成, 简单方便且检测精度高, 其主要包含摄像头检测模块、食品识别与分析模块、语音提醒及结果可视化展示模块, 主要工作流程如图 1 所示如下:

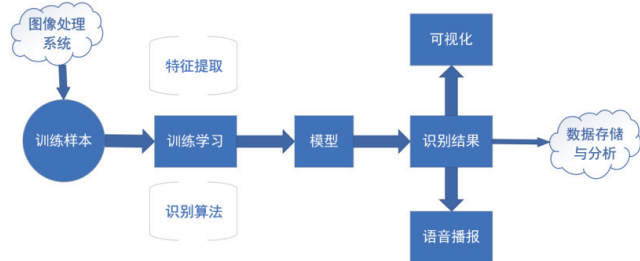


图1 系统流程图

目标捕获: 通过摄像头拍摄识别范围内的图片和视频作为原始输入, 经过预处理后传递给目标识别模块。

目标识别: 该模块利用多级卷积神经网络训练的初始模型进行识别, 并在实际工作场景中利用摄像头拍摄的照片和视频迭代学习, 更新识别模型, 不断提高识别正确率。

检测结果语音播报、可视化与存储: 在违禁物品识别后即可将识别结果信息通过语音合成技术播报提醒和可视化展示, 并将识别信息上传到学校信息中心云存储平台, 达到违禁食品的及时预警和监督的目的。

(一) 图像处理系统

系统高精度的识别目标领域的对象依赖于大量的领域样本作为输入对模型进行训练, 然而现实问题是样本往往是不充足和不完备的, 比如本系统处理对象是食品, 包含零食和早、中和晚餐, 种类繁多, 期望人工拍摄每一种食品的图像作为模型的训练素材是不现实的, 因此就需要一个定制的图像处理系统来解决这个问题, 其主要功能是:

1. 图像采集。根据识别目标类型设计图像采集爬虫程序, 从公开的互联网上爬取满足要求的图片。
2. 图像预处理。图像预处理主要实现图像的减法, 即对采集的图像进行规范化操作, 过滤掉不清晰的、图片格式、大小等不满足要求的图片, 确保训练样本图片可用性。
3. 图像增强。图像增强主要功能是对图片做加法, 即基于规范化后的图片采用图像增强技术增加图片的数量, 避免因图像数量稀少和场景单一造成模型训练不足而产生过拟合现象, 我们采用的增强技术包括图像灰度调整、图像旋转、缩放和剪裁。

本系统经过定制 python 爬虫采集程序爬取了校园常见零食图片 3000 张, 早餐类图片包子 2550 张、鸡蛋 2020 张和油条 900 张, 并采用基于 Anaconda3 集成开发工具, 使用 Python 语言的 OpenCV 相关图像处理库进行图形处理系统的开发, 经过图像处理系统处理后, 最终扩充到近 100000 张图片作为训练样本集

(二) 多层卷积神经网络集成模型框架

本系统设计了基于深度学习的多层卷积神经网络集成模型框架进行食品的视觉图像识别和识别结果的语音合成播报。集成模型框架如图 2 所示:



图2 卷积神经网络集成模型框架

首先, 从图像处理系统获取增强和规范化后的大量图片作为样本数据, 将图片按一定比例分为训练图片集和测试图片集。

其次, 通过训练图片集迭代训练出一个食品识别模型 (这个模型结构即基于深度学习的多层卷积神经网络结构, 模型结构具体在下文介绍), 并用测试图片集验证该模型识别效果, 这个过

程涉及多次测试和模型参数的迭代优化。

最后,使用优化后的模型对待验证的新目标图片进行识别,评估模型的可用性,模型确定达到使用标准后即可集成到系统中被调用。

上文述及的基于深度学习的多层卷积神经网络模型由数据输入层、卷积层、激励层、池化层、全连接层以及输出层组成。其作用是识别这张图片里具体有哪些目标、分别是什么目标,例如针对一张图片,识别图中是包子、鸡蛋还是米饭,是谁在拿着这些食品。

1. 输入层。卷积神经网络的输入层主要是读入原始图像将其输入特征进行标准化处理,本系统主要的采用的处理方法包括灰度化、去均值,归一化和PCA/白化。

2. 卷积层。卷积层是卷积神经网络最重要的一个层次,其功能是对输入的数据进行特征提取,内部包含有多个卷积核,一个卷积核学习一个特征,且我们的卷积操作是逐层设计的,因为第一层卷积层一般只能提取一些低级的特征,而更高的卷积层能从低级特征中迭代提取更复杂的特征,这也是符合我们人类生物认知规律的。

结合卷积层的理论原理和食品领域特征,我们设计了4层卷积操作,每层卷积核的个数倍次增加,即第一、二、三、四层分别设计了32、64、128和256个卷积核,以便逐层深入学习更细粒度的特征,卷积核的大小都是3*3,即采用参数共享方式,其科学依据是食品图片的局部性原理,即相邻部分的图像统计特性大体是类似的,所以对于这个图像上的所有位置都使用同样的学习特征。

3. 激励层。其作用是把卷积层的输出结果做非线性映射,这里我们采用ReLU(修正线性单元)激励函数,梯度简单而且收敛快,同时因为系统检测对象的像素值都在0~255范围内,不存在特征值小于0时,避免了反向传播算法收敛速度慢的问题。

4. 池化层。池化层位于连续的卷积层中间,其作用是将输入图像进行缩小,即过滤无关紧要的信息来压缩图像,使其保留的特征是最能够表达图像信息的。具体策略上我们设计使用maxpooling来进行池化操作,在原始图像的数据层上划分一个个小块,在每个小块中选择该块像素最大值代表这个小块中所有的值放到下一层,其理论依据是图像的局部性原理,在一幅图中,每个像素点和其周边的点大致是一样的,所以用一个最大值点代替其周边点是可行的。

5. 全连接层。该层常其位于网络的尾部并且两层之间的所有神经元都是拥有权重连接的,对特征进行重新拟合,在整个网络中起到“分类器”的作用,本系统设计了两层全连接层,输出节点数分别为256和128,逐步规约分类目标。

6. 输出层。输出层是卷积神经网络的最后一层,做好最后目标结果的输出,本系统设计使用softmax函数来对目标食品进行识别,输出分类标签,本层输出节点32,分别表示系统设计要识别的32种违禁食品。

(三) 识别结果处理模块

识别结果处理模块主要针对上文网络模型的识别结果进行后续处理,包括如下3个子模块,如图3所示:



图3 识别结果模块流程图

1. 自然语言合成播报提醒模块。其主要功能是根据模型的识别结果合成语音播报提醒,达到即时阻止学生将违禁食品带进受限区域的目的,主要包括如下几个步骤:

(1) 设计文本提醒模板。预定义提醒文本的模板配置规则,如样例: \${person_name} 同学,你好,请不要将 \${food_name} 带到 \${location},谢谢配合。

(2) 生成提醒文本。根据网络模型的识别结果,替换配置模板中的\$标识的变量,完成提醒文本的生成。

(3) 合成语音提醒播报。基于Anaconda集成开发平台,使用Python语言的pyttsx3模块完成从文字到语音的转换,pyttsx3模块相比其他语音转换模块的优势是更具通用性,内置语音引擎,可以自由设置语音的音量、频率、变声等,也可以格局需要循环播放,设置提醒次数等定制需求。

2. 识别结果存储与分析模块。其主要功能是将上文的识别原始结果存储到校园信息中心数据库系统,并设计周期性任务进行离线数据分析,结合校园既有信息系统进行数据的补充完善,为校园的智慧化管理提供决策依据。

3. 可视化展示模块。其主要功能是针对模块2的分析结构进行可视化展示,便于直观形象的了解整个校园的违禁纪录的发生情况。

二、系统验证

(一) 网络模型迭代训练

网络模型训练共采用了90000张不同的食品种类图片训练,其中训练图片占八成,剩下的两成进行测试,在训练和测试时将图片进行随机打乱,模型使用这些图片进行了10次的迭代训练和测试,训练结果显示损失值为0.05左右,准确率在95%以上;测试结果的损失值为0.0325,准确率为96%以上,达到了可用的要求。

(二) 模型验证

对基于大量训练集训练后的模型进行新图片的泛化测试,采用了10000多张图片,为了评估模型对非训练集中图片的识别精度,测试集中的图片和训练集中的图片一样需要打上标签。根据测试集的准确率数据图可以看出数据在测试之初从低往上增长,增长到95%以后趋于平缓,说明系统采集了大量不同种类的样本数据进行迭代训练对提升模型识别精度是有很大的作用的。

三、结语

应用卷积神经网络结构进行计算机视觉识别是目前机器学习领域的一个主流的研究和应用方向,随着机器学习的飞速发展,深度学习技术也越来越先进和成熟,人们对于人工智能的热情也日渐增强。而机器视觉是通过摄像头将摄取到的目标转化成图像信号并进行目标的识别与分析,并对分析结果进行合成语音播报,相当于让机器长了眼睛和嘴巴,能“看”会“说”,越来越接近人的智能。本文进行深度学习技术应用于食品识别的横向研究,选择了在机器视觉领域具有十分优秀应用效果的卷积神经网络应用到我们的“食品目标识别和监督系统”项目,并经过系统的测试,达到了较好的识别精度,成功地完成了课题的研究任务。

参考文献:

- [1] 王慧玲, 蔡小龙, 武港山. 基于深度卷积神经网络的目标检测技术的研究进展[J]. 计算机科学, 2018, 45(9): 11-19.
- [2] 何小嵩, 张占文, 荣伟彬. 基于计算机视觉的微粒缺陷检测及分类方法[J]. 强激光与粒子束, 2017, 29(8): 75-79.