

基于半监督深度卷积生成对抗网络的注塑瓶表面缺陷检测模型



谢源 苗玉彬 许凤麟 张铭

上海交通大学机械与动力工程学院 上海 200240

(xieyuansjtu@163.com)

摘要 注塑瓶表面缺陷检测是注塑成型工艺流程中的重要环节,但生产中存在缺陷的注塑瓶样本数量相对匮乏,使得应用深度学习算法进行缺陷检测时容易产生过拟合现象。针对上述问题,文中提出并构建一种半监督(Semi-supervised)深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Network, DCGAN)模型。该模型首先使用 HSV(Hue Saturation Value)颜色空间转换与大津算法(Otsu)对原始注塑瓶图像进行预处理得到训练集;然后组合学习任务,使得 DCGAN 的无监督判别器与注塑瓶表面缺陷检测的监督分类器共享卷积层参数,同时修改损失函数,在 DCGAN 模型的 Wasserstein 距离中加入交叉熵;最后使用 Adam 优化器进行模型训练。实验结果表明,该模型能够准确分辨具有缺陷的注塑瓶样本,分类准确率达到 98.65%。与传统的机器学习算法以及采用数据增强的卷积神经网络模型相比,所提模型的分类准确率更高,且较好地避免了过拟合现象,能满足注塑瓶生产中表面缺陷的自动检测需求。

关键词: 深度卷积生成对抗网络;半监督;小样本;缺陷检测;注塑瓶

中图法分类号 TP277

Injection-molded Bottle Defect Detection Using Semi-supervised Deep Convolutional Generative Adversarial Network

XIE Yuan, MIAO Yu-bin, XU Feng-lin and ZHANG Ming

School of Mechanical Engineering, Shanghai Jiaotong University, Shanghai 200240, China

Abstract Defect detection of injection-molded bottles is an important part of injection molding. Due to the relatively few defective samples in production, the model tends to over-fit when using deep learning algorithm. In order to solve this problem, a defect detection model based on semi-supervised deep convolutional generative adversarial network (DCGAN) is proposed. Firstly, the model preprocesses the original images using HSV color space transformation and Otsu threshold segmentation methods. Then, the learning tasks are combined so that the unsupervised discriminator and the supervised classifier share convolutional parameters. At the same time, the loss function is modified, which consists of cross entropy and Wasserstein distance. Finally, the model is fine-tuned using Adam optimizer. The experimental results show that the model can distinguish the defective samples, achieving an accuracy of 98.65%. Compared with traditional machine learning algorithm and CNN model with data augmentation, the proposed model avoids over-fitting.

Keywords Deep convolutional generative adversarial network, Semi-supervised, Small-sized datasets, Defect detection, Injection-molded bottle

1 引言

注塑成型是一种重要的塑料制品加工工艺。生产过程中,受原材料物理化学性质变化^[1]、注塑机设备运行参数及状态变化^[2]、自动化控制系统精度^[3]等因素的影响,塑料制品不可避免地会产生制造缺陷,而诸如化妆品等使用注塑瓶包装的消费产品对外观具有较高要求,因而注塑瓶表面缺陷检测在实际生产中具有重要意义。

目前,生产中的注塑瓶表面缺陷检测主要依靠人工在线检测和离线抽检相结合的方式,需要花费大量人力成本,并且

检测效率不高。为了提高生产的自动化程度,研究者们开展了基于机器视觉的表面缺陷检测方法的研究,如 Lv 等^[4]研究了针对注塑制品的多种视觉特征提取方法。机器视觉检测方法的研究重点为特征提取,须根据不同类型的特征设计处理流程。为提高算法的通用性,基于卷积神经网络(Convolutional Neural Network, CNN)的缺陷检测模型逐渐受到关注,与基于传统机器视觉方法的人工特征提取相比, CNN 能够自动提取图像中的特征^[5],且特征更具区分度^[6]。但多层神经网络模型需要大量数据进行训练,而生产过程中注塑瓶的良品率较高,存在缺陷的样本数量有限,使用小样本集训练复杂

收稿日期:2019-07-13 返修日期:2019-11-04 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(51975361)

This work was supported by the National Natural Science Foundation of China (51975361).

通信作者:苗玉彬(ybmiao@sjtu.edu.cn)

神经网络通常会产生过拟合现象,导致模型泛化能力减弱,因此需要增加现有数据的数量与差异性以改善分类准确率。针对上述问题,研究者们对数据增强算法进行了探索。Bjerrum等^[7]提出使用仿射变换生成新样本,并与真实样本混合作为训练集,提高了模型的鲁棒性。但是CNN对旋转、平移具有不变性,因此该方法对准确率的提升有限。数据增强算法也可基于生成模型来实现^[8-9],如Antoniou等^[10]提出用条件生成对抗网络(Conditional Generative Adversarial Network, cGAN)生成不同肤色、表情的人脸图像进行数据增强。与仿射式图像变换相比,生成样本更符合人的主观认知,后续分类模型的准确率也更高。Shin等^[11]也利用生成对抗网络模型(Generative Adversarial Network, GAN)解决了医疗图像中样本分布不均匀的问题,但由于模型容量限制,GAN无法重构出复杂细节特征,即生成数据分布与真实数据分布并不完全重合。在GAN模型的基础上,Radford等^[12]使用反卷积神经网络作为生成器,使用卷积神经网络作为判别器,提出了无监督的深度卷积生成对抗网络(Deep Convolutional Generative Adversarial Network, DCGAN),使得GAN在图形、语音领域内应用得更加广泛。

应用上述数据增强的方法可以生成大量无标签注塑瓶样本,并通过人工标记进行缺陷检测分类器的训练。由于生成数据分布与真实数据分布并不是完全相同的,因此该方法对分类准确率的提升有限。针对生成模型生成的无标签样本,半监督学习^[13]理论提供了一种新的应用思路,即模型同时利用无标签样本与有标签样本进行训练。与纯监督模型相比,半监督模型能够利用额外的无标签样本优化分类超曲面。Lasserere等^[14]提出当共享无监督真实数据生成模型与监督分类器的参数时,通过权衡无监督损失项与监督损失项,可以得到比纯监督模型更好的效果。Salakhutdinov等^[15]也描述了在学习核函数过程中使用真实数据生成模型优化分类器的方法。但是,半监督模型中使用的无标签样本都采样自真实数据分布,而生产中无法直接获得注塑瓶图像样本的概率分布模型。

本文针对注塑瓶缺陷检测时因样本数量过少导致的多层神经网络模型过拟合的问题,借鉴DCGAN与半监督学习的相关研究进展,提出了一种半监督DCGAN模型。该模型首先利用HSV颜色空间变换与Otsu阈值分割对图像数据进行预处理,从而得到训练集;然后在DCGAN模型的无监督判别器中加入监督分类器,使判别器与分类器共享底层卷积参数;并修改损失函数,在Wasserstein距离中加入交叉熵,使得训练过程中监督学习与无监督学习能够同时进行;最后利用Adam优化器对模型进行训练。

2 基于半监督深度卷积生成对抗网络的模型

2.1 DCGAN模型

DCGAN模型是一个无监督的生成模型,典型的DCGAN模型分为生成器 G 与判别器 D ,二者均为CNN神经网络,如图1所示。其中,生成器拟合真实数据分布,将噪声映射至样本空间;判别器则判断输入样本的真伪。通过交替训练生成器与判别器,生成器输出的数据分布将逐渐逼近真实分布。

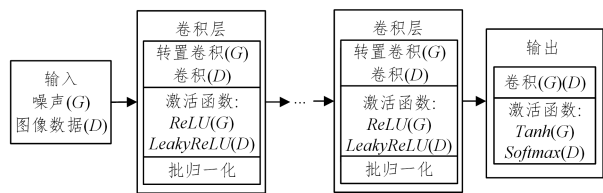


图1 DCGAN结构

Fig.1 Structure of deep convolutional generative adversarial network

相比标准的CNN神经网络,DCGAN生成器与判别器主要有以下改变:(1)取消池化层,用带步幅的卷积层实现降采样功能。(2)取消生成器中的全连接层,改为全卷积网络,加快模型的收敛速度。(3)对隐层输出使用批归一化,避免梯度消失或梯度爆炸问题。(4)生成器的隐层采用ReLU激活函数,输出层采用Tanh激活函数;判别器的隐层采用LeakyReLU激活函数,输出层采用Softmax激活函数。

DCGAN模型利用卷积层的特征表示能力,能够生成具备精细特征的注塑瓶样本。作为生成模型,DCGAN可替换半监督模型中的真实数据生成器,提供训练所需的无标签样本。

2.2 改进的半监督DCGAN模型

监督分类模型的优化目标通常为标签与模型预测结果之间的交叉熵,而无监督DCGAN模型的优化目标为两种数据分布之间的散度,本文将两者按照下述方式组合从而得到半监督DCGAN模型。

(1)监督分类器与DCGAN模型的无监督判别器共用同一输入层;监督分类器与无监督判别器都以注塑瓶图像为输入,两者具有相同的输入空间,因此可共用输入层。

(2)监督分类器与DCGAN模型的无监督判别器共享底层卷积参数;无监督判别器的底层为卷积层,而基于CNN的缺陷检测分类器也采用卷积为特征提取工具。卷积层提取的注塑瓶纹理、边缘等特征^[6]可同时用于判断样本真伪与是否存在缺陷,因此两者可共享卷积层。

基于上述分析,本文拓展无监督判别器的输出维度,增加监督分类输出维,形成改进的半监督DCGAN模型,其结构如图2(a)所示。生成器保持不变,仍然从噪声中采样并生成无标签样本,其结构如图2(b)所示。对于有 k 个类别的监督分类问题,半监督判别器的输出维度为 $k+1$ 维。其中,前 k 维表示真实样本所属类别,第 $k+1$ 维判断输入样本的真伪,如图2(c)所示。同时,本文在模型优化目标中加入监督损失项,同步进行无监督学习与监督学习。

生产中仅需判别是否存在缺陷,属于二分类问题,因此判别器的输出维度为3。无监督学习采用带梯度罚项的Wasserstein距离^[16]作为损失函数,监督学习采用交叉熵作为损失函数,建立的损失函数如下:

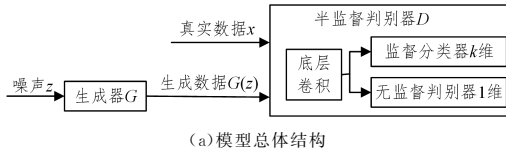
$$L_D = L_{ms}^w + L_s + L_{gp} \\ = -E_{x \sim p_{data}} [D(x)]|_{y=k+1} + E_{z \sim p_z} [D(G(z))]|_{y=k+1} - \eta \\ E_{x, y \sim p_{data}} (x, y) \log [D(x)]|_{y=k+1} + \lambda E_{z \sim p_z} \\ [(\|\nabla D(x)\|_2^2 - 1)^2] \quad (1)$$

$$L_G = -E_{z \sim p_z} [D(G(z))]|_{y=k+1} \quad (2)$$

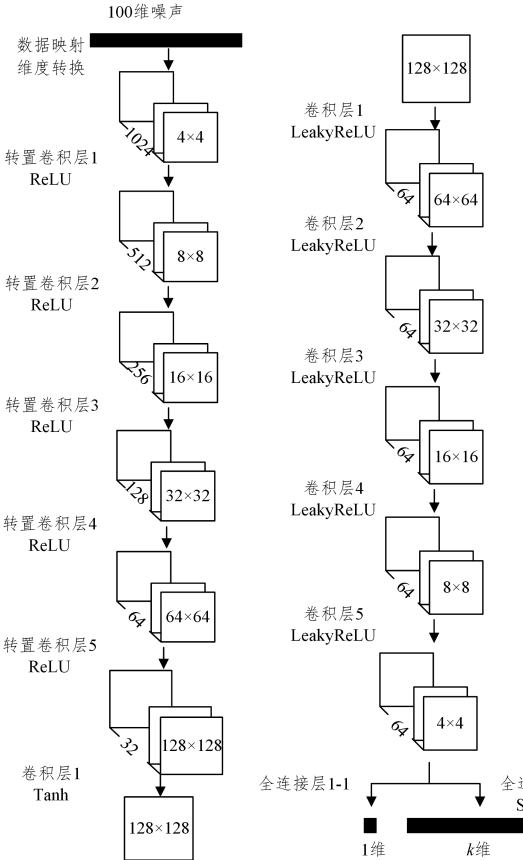
其中, L_D 表示判别器的损失函数,由监督损失项 L_s 、无监督损失项 L_{ms}^w 和梯度惩罚项 L_{gp} 组成; L_G 表示生成器的损失函数; η

和 λ 为权重系数,而 Wasserstein 距离项的权重系数默认为 1。

在一轮训练过程中,生成器参数更新规则与 DCGAN 模型相同。针对判别器的参数更新,输入无标签样本时,损失函数为带梯度罚项的 Wasserstein 距离;输入有标签样本时,损失函数中加入交叉熵来指导监督学习。超参数 η 表征在参数更新过程中交叉熵相比于 Wasserstein 距离的重要程度。



(a) 模型总体结构



(b) 生成器 G 的结构

(c) 半监督判别器 D 的结构

图 2 半监督 DCGAN 模型的结构

Fig. 2 Structure of semi-supervised deep convolutional generative adversarial network

2.3 无标签样本提供的弱先验知识

半监督 DCGAN 模型中,无标签样本不会直接影响监督损失项的计算,但损失函数中无监督损失项会指导模型中共享参数的更新,从而对缺陷检测分类结果产生影响。半监督学习认为^[12],在输入空间中紧密聚集的样本应该被映射到类似的表示。本文中生成器输出的数据分布与真实数据分布不完全重叠,但优化过程中两个分布的散度逐渐减小,可以认为两者是相似的。当有标签样本不足时,无标签样本可对最终得到的分类超曲面进行优化,得到比纯监督学习更好的泛化能力。

此外,本文的判别器同时判别样本真伪与是否存在缺陷,属于多任务学习模型,模型对参数施加软约束,使得参数更新倾向于使卷积层提取出两种任务共同使用的特征^[17]。由于后续在注塑瓶缺陷样本预处理过程中消除了背景干扰,底层卷

积提取的特征主要包括注塑瓶的边缘和纹理等特征,该种特征可用于判别样本的真伪,减小散度,也可用于判别样本是否存在缺陷,减小交叉熵。相比分别构建生成模型与分类模型的方法,多任务学习的参数冗余程度更弱,能够改善过拟合现象。

与数据增强算法中的人工指定标签相比,半监督模型具有更弱的先验条件。该模型仅指定输入空间中紧密聚集的样本应该被映射至相似的类别,卷积层提取的特征可被判别器与分类器共享,因此其在训练过程中能缓解过拟合现象,具备更强的泛化能力。同时,判别器与分类器同步训练,避免了人工标记数据所需的庞大工作量。

3 实验与分析

3.1 优化器设置与超参数选择

使用 Adam 优化器^[18]进行模型训练,优化器参数为:矩估计指数衰减速率 $\beta_1 = 0.9, \beta_2 = 0.98$;步长 $\epsilon = 10^{-3}$;数值稳定参数 $\delta = 10^{-9}$;有偏一阶矩与有偏二阶矩的初始值 $s_0 = 0, r_0 = 0$ 。设模型参数为 θ ,则在一轮训练中 θ 的更新规则如下。

(1) 计算 t 时间步的损失函数的梯度:

$$g_t = \nabla_{\theta} L(\theta_{t-1}) \quad (3)$$

(2) 更新有偏一阶矩与有偏二阶矩:

$$s_t = \beta_1 s_{t-1} + (1 - \beta_1) g_t \quad (4)$$

$$r_t = \beta_2 r_{t-1} + (1 - \beta_2) g_t \odot g_t \quad (5)$$

(3) 修正有偏一阶矩与有偏二阶矩:

$$\hat{s}_t = \frac{s_t}{1 - \beta_1^t} \quad (6)$$

$$\hat{r}_t = \frac{r_t}{1 - \beta_2^t} \quad (7)$$

其中, β_1^t, β_2^t 分别表示 β_1, β_2 的 t 次方。

(4) 更新模型参数 θ :

$$\theta_t = \theta_{t-1} - \epsilon \frac{\hat{s}_t}{\sqrt{\hat{r}_t + \delta}} \quad (8)$$

取模型超参数 $\eta = 0.2, \lambda = 0.5$,由于 η 是监督学习权重系数,因此上述超参数设置表明在实际训练过程中,开始阶段模型更关注生成器的数据拟合表现;而当训练过程中 Wasserstein 距离逐渐减小时,交叉熵在损失函数中开始占主导地位,模型的分类准确率将迅速提高。

3.2 训练过程

图 3 为训练过程中的损失函数变化曲线,图 4 为不同迭代次数下的生成样本,随着迭代次数的增加,生成样本逐渐具备精细特征。

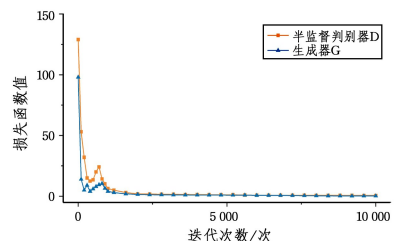


图 3 训练过程中损失函数的变化

Fig. 3 Variations of loss functions during training

迭代 100 次时,生成样本仅产生圆形特征;迭代 1 000 次时,生成样本表现出注塑瓶底边缘特征及光斑;迭代

5000次时,生成样本具备完整的细节特征与缺陷特征,但部分样本拟合情况较差;迭代10000次时,生成样本边缘、纹理

等特征与真实样本基本相同,无法主观判别两类样本,生成模型已较好地拟合了真实数据分布。

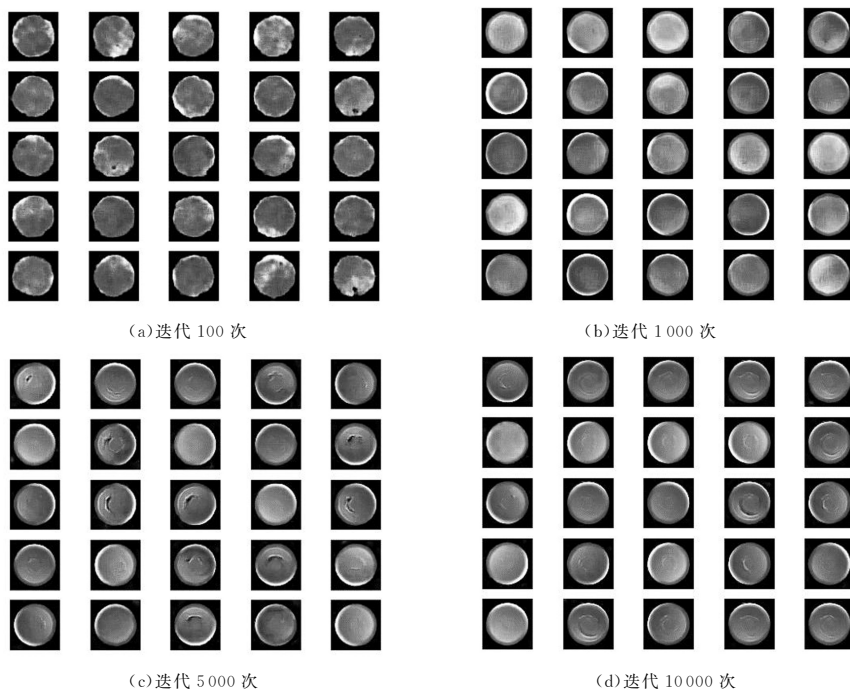


图4 不同迭代次数下的生成样本实例

Fig. 4 Generated images under different iterations

使用测试集对不同迭代次数下的判别器进行验证,如表1所列。可见,随着迭代次数的增加,分类器准确率逐步提升。模型完成训练时,半监督 DCGAN 模型中分类器的准确率达到 0.9865。

表1 不同迭代次数下评价指标的变化

Table 1 Variations of evaluating results under different iterations

迭代次数	Wasserstein 距离	准确率	精确率	召回率	F1 分数
100	-3.0963	0.7043	0.5789	0.5641	0.5714
1000	4.7309	0.8674	0.8201	0.7949	0.8073
5000	0.5721	0.9493	0.9302	0.9231	0.9266
10000	0.3695	0.9865	0.9845	0.9769	0.9807

3.3 模型效果对比

为验证本文模型的效果提升,将本文模型分别与基于主成分分析(Principal Component Analysis, PCA)和反向传播(Backpropagation, BP)神经网络的缺陷检测模型、基于尺度特征不变(Scale-invariant feature transform, SIFT)和支持向量机(Support Vector Machine, SVM)的缺陷检测模型、基于仿射数据增强的 CNN 网络以及基于生成数据增强的 CNN 网络进行比较,结果如表2所列。

表2 不同模型的效果比较

Table 2 Evaluations of different models

模型	准确率	精确率	召回率	F1 分数	运算时间/ms
PCA+BP 神经网络	0.7688	0.6755	0.6513	0.6632	13.932(CPU)
SIFT+SVM	0.8554	0.8173	0.8692	0.8425	6.956(CPU)
CNN(仿射数据增强)	0.9358	0.9192	0.9481	0.9334	2.793(GPU)
CNN(生成数据增强)	0.9541	0.9275	0.9612	0.9440	2.904(GPU)

表2中,准确率、精确率、召回率、F1分数等评价指标^[19]的定义如下:

$$R_{\text{accuracy}} = \frac{N_{tp} + N_{tn}}{N_{tp} + N_{fp} + N_{tn} + N_{fn}} \quad (9)$$

$$R_{\text{precision}} = \frac{N_{tp}}{N_{tp} + N_{fp}} \quad (10)$$

$$R_{\text{recall}} = \frac{N_{tp}}{N_{tp} + N_{fn}} \quad (11)$$

$$R_{F1} = \frac{2 \times R_{\text{precision}} \times R_{\text{recall}}}{R_{\text{precision}} + R_{\text{recall}}} \quad (12)$$

其中, N_{tp} 为存在缺陷的样本被模型分类正确的次数, N_{tn} 为无缺陷样本被模型分类正确的次数, N_{fp} 为无缺陷样本被模型误分类的次数, N_{fn} 为存在缺陷的样本被模型误分类的次数。

3.4 实验结果及分析

由表3可见,本文模型与BP网络和SVM分类器相比,存在如下优势:

(1)BP网络与SVM分类器能够提取到注塑瓶样本中少量的缺陷特征,分类准确率分别为76.88%和85.54%,均高于随机选择。但由于模型拓扑结构简单,不具备较强的数据拟合能力,出现了较多的误分类。

(2)基于深度学习的模型具有更高的准确率,这说明与BP网络中的全连接层以及SVM分类器中的统计学习方法相比,卷积层在处理图像数据、自动提取图像特征方面的性能更优越,也证明了深度学习算法在注塑瓶缺陷检测领域应用的必要性。

(3)传统的机器学习模型无GPU算力支持。在GPU加速下,本文模型的前向运行时间优于传统模型,尽管本文模型的拓扑网络结构更为复杂,但是单次检测的效率却更高。

本文模型与两种采用数据增强方法的CNN模型相比,存在如下优势:

(1)在准确率方面,相比仿射数据增强的 CNN 模型与生成数据增强 CNN 模型,本文模型的准确率分别提升了 5.07% 和 3.24%。

(2)在精确率方面,两种数据增强的 CNN 模型的精确率偏低。由精确率公式可知, N_{fp} 较大表明模型对无缺陷样本的误分类率较高,数据增强过程中新样本产生了真实样本不具备的特征,且 CNN 模型提取了该类特征并将其作为判断样本存在缺陷的依据,导致过拟合。

(3)在召回率方面,两种数据增强 CNN 模型的召回率偏高。由召回率的公式可知, N_{fn} 较小表明模型倾向于将输入样本判断为存在缺陷,印证了关于精确率的分析。

综上,在处理注塑瓶缺陷检测问题时,本文模型相比传统的机器学习模型具有更强的泛化能力,相比数据增强的 CNN 模型能够较好地避免过拟合现象,准确率更高。

结束语 注塑瓶的表面缺陷检测是自动化注塑工艺中的重要环节。本文针对注塑瓶缺陷检测时因样本数量少导致的多层神经网络模型过拟合问题,提出并构建了一种半监督深度卷积生成对抗网络模型。该模型在 DCGAN 模型的判别器输出中增加分类器的输出维度,合并多个学习任务,减弱了参数冗余程度以及模型优化的难度;并通过带梯度罚项的 Wasserstein 距离损失函数指导无监督学习,提高了卷积层的特征提取能力;同时将卷积网络参数共享至缺陷检测监督分类器,通过有标签样本优化分类准确率,提高了小样本下分类器的泛化能力。实验验证表明,与基于传统 PCA 特征和 BP 神经网络的分类器,以及基于 SIFT 特征和 SVM 的分类器相比,本文模型在分类准确率上分别提升了 21.97% 和 13.11%;与基于仿射数据增强的 CNN 模型以及生成数据增强的 CNN 模型相比,本文模型在分类准确率上分别提升了 5.07% 和 3.24%,达到了 98.65%,且精确率与召回率无明显上下波动,有效避免了过拟合现象,能够满足注塑瓶生产中表面缺陷的自动检测需求。

参 考 文 献

- [1] GUO Z Y, LI D Q. Study on Deformation of Warping Fermentation starter of Injection Molded Products[J]. Plastic Technology, 2001(1): 22-24.
- [2] REN G C, MIAO X Q, GUO Z G. Optimum design of the double-toggle clamping unit for injection molding machine[J]. Machine Design and Manufacture, 2009(1): 18-19.
- [3] ZHANG H J, QUAN L, LI B. Comparative Study on Energy Efficiency of the Electro-hydraulic Control System in Injection Molding Machine[J]. Chinese Journal of Mechanical Engineering, 2012, 48(8): 180-187.
- [4] LV Z. Research on Key Techniques of Vision Inspection for Injection Molding Products[D]. Massachusetts: Northeastern University, 2009.
- [5] ATHA D J, JAHANSHAHI M R. Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection[J]. Structural Health Monitoring, 2018, 17(5): 1110-1128.
- [6] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks[C]// European Conference on Computer Vision. Springer, Cham, 2014: 818-833.
- [7] BJERRUM E J. SMILES enumeration as data augmentation for neural network modeling of molecules[J]. arXiv: 1703. 07076, 2017.
- [8] CHEN W B, GUAN Z X, CHEN Y J. Data augmentation method based on conditional generative adversarial net model[J]. Journal of Computer Applications, 2018, 38(11): 3305-3311.
- [9] CHENG X Y, XIE L, ZHU J X, et al. Review of Generative Adversarial Network[J]. Computer Science, 2019, 46(3): 74-81.
- [10] ANTONIOU A, STORKEY A, EDWARDS H. Data Augmentation Generative Adversarial Networks[J]. arXiv: 1711. 04340.
- [11] SHIN H C, TENENHOLTZ N A, ROGERS J K, et al. Medical Image Synthesis for Data Augmentation and Anonymization using Generative Adversarial Networks[J]. arXiv: 1807. 10225.
- [12] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. arXiv: 1511. 06434, 2015.
- [13] LI T, YING N, YU X, et al. Semi-supervised learning in unbalanced and heterogeneous networks[J]. arXiv: 1901. 01696.
- [14] LASSERRE J A, BISHOP C M, MINKA T P. Principled Hybrids of Generative and Discriminative Models[C]// IEEE Computer Society Conference on Computer Vision & Pattern Recognition. 2006.
- [15] SALAKHUTDINOV R, HINTON G E. Learning a Nonlinear Embedding by Preserving Class Neighbourhood Structure[J]. Journal of Machine Learning Research, 2007, 2: 412-419.
- [16] GULRAJANI I, AHMED F, ARJOVSKY M, et al. Improved Training Of Wasserstein Gans[C]// Advances in Neural Information Processing Systems. 2017: 5767-5777.
- [17] GOODFELLOW I, BENGIO Y, COURVILLE A, et al. Deep learning[M]. Cambridge: MIT Press, 2016.
- [18] KINGMA D P, BA J. Adam: A method for stochastic optimization[J]. arXiv: 1412. 6980, 2014.
- [19] MAO Y C, WANG J, CHEN X L, et al. Dam Defect Recognition and Classification Based on Feature Combination and CNN[J]. Computer Science, 2019, 46(3): 267-276.



XIE Yuan, born in 1997, postgraduate. His main research interests include machine learning and control algorithms.



MIAO Yu-bin, born in 1973, Ph.D, associate professor. His main interests include intelligent devices, smart sensors, and 3-dimensional image processing.