

文章编号: 2095-2163(2021)11-0157-05

中图分类号: TP181;R318

文献标志码: A

# 基于 EfficientNet 的胶囊内镜图像多病灶协同检测

刘 张, 郭旭东, 张璐璐

(上海理工大学 医疗器械与食品学院 上海 200093)

**摘要:** 消化道息肉和慢性炎症性疾病的进一步发展容易引发癌变,深度学习的方法能够辅助医生自动检查胶囊内镜图像中的病灶,实现医学图像分类,目前的自动检测研究模型很少权衡模型的准确度与速度,难以适应临床中对多病灶协同检测的需求。为此,本文提出一种基于轻量级网络的迁移学习方法,以辅助检测消化道常见病变。经与 ResNet50 和 InceptionV4 网络进行对比,在公开的 CAD-CAP 数据集上的实验结果表明,该方法对消化道血管畸形病变的敏感度达 94.74%,炎症性病变的敏感度达 89.76%,模型准确率为 94.72%,平均 AUC 值为 0.94,检测速度为 73 帧/秒,所提出的方法在准确率和运行效率上均优于传统方法,能够为消化道疾病的辅助诊断提供有价值的帮助。

**关键词:** 医学图像分类; 胶囊内镜; 迁移学习; 轻量级网络

## Multiple lesions detection in capsule endoscopic images based on EfficientNet

LIU Zhang, GUO Xudong, ZHANG Lulu

(School of Medical Instrument and Food Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

**[Abstract]** The development of digestive tract polyps and chronic inflammatory diseases may lead to carcinogenesis. Deep learning method can assist doctors to automatically examine lesions in capsule endoscopic images. The current automatic detection research model rarely weighs the accuracy and speed of the model, which is difficult to adapt the needs of multiple lesion detection in clinical practice. Therefore, a migration learning method based on lightweight network is proposed to classify common diseases of the digestive tract. Finally, compare with the Resnet50 and InceptionV4 networks. The results of CAD-CAP data set showed that the precision for vascular lesions was 94.74% and that for inflammatory lesions was 89.76%. The accuracy was 94.72%. The average AUC value was 0.94, and the detection speed was 73 frames/s. The proposed method is superior to the traditional methods in accuracy and speed, which can provide valuable help for the auxiliary diagnosis of digestive tract diseases.

**[Key words]** medical image classification; capsule endoscopy; transfer learning; lightweight network

## 0 引言

消化道癌症的前身主要是息肉和一些慢性炎症性疾病,因此消化道疾病的早检早治对于降低癌变、死亡率有重大意义<sup>[1]</sup>。无线胶囊内窥镜(Wireless Capsule Endoscopy, WCE)作为一种安全无创的消化道检查方式,解决了传统内镜设备对于小肠和大肠部分位置不能检测的问题,但一次完整的检查拍摄过程用时约 8 h,平均获得 60 000 张图像,而包含病灶的仅占很少一部分<sup>[2]</sup>。人工阅片的过程费时费力,且长时间的人工阅片过程容易增加漏诊率,因此自动的消化道病灶检测方法的研究显得尤为重要。

传统的机器学习方法通过人工提取图像中的颜色特征、形状特征和纹理特征,使用支持向量机等分类器实现内镜图像中病灶的自动检测。近年来,深

度学习作为人工智能领域极具应用前景的机器学习算法,被广泛应用于医学影像的辅助诊断中<sup>[3]</sup>。为了适应临床应用需求,研究人员已经尝试开展内镜图像多病灶协同检测的研究,如:基于 ResNet 网络利用滑动块逐行移动的方法识别乳糜泻、正常组织和非特异性十二指肠炎<sup>[4]</sup>;结合 ResNet34 和 Faster-RCNN,通过迁移学习的方式区分 6 种胃囊内窥镜图像,利用建立的大型内镜数据集证明了多病灶检测方法的临床应用可行性<sup>[5]</sup>;通过迁移学习 DenseNet 模型分类检测胶囊内镜图像中的 37 种病变,发现不同类别之间的相似性会对检测精度产生影响<sup>[6]</sup>;使用微调 CNN 的方法区分胃镜图像中的正常、良性溃疡和癌症<sup>[7]</sup>;修改 AlexNet 的多路输出为两路输出,表示正常和异常,成功检测出溃疡和糜烂病灶<sup>[8]</sup>;利用 RetinaNet 网络结合 SSD 网络体系结

**基金项目:** 国家自然科学基金(61001164);上海市自然科学基金(15ZR1428200)。

**作者简介:** 刘 张(1997-),男,硕士研究生,主要研究方向:医学图像处理和生物医学图像人工智能研究;郭旭东(1980-),女,博士,副教授,硕士生导师,主要研究方向:精密医疗器械和医学图像智能辅助诊断研究;张璐璐(1994-),女,硕士,工程师,主要研究方向:医学图像处理、人工智能辅助诊断研究。

**通讯作者:** 郭旭东 Email: guoxd@usst.edu.cn

收稿日期: 2021-08-23

构,分别对内窥镜图像中的 4 种类型的病变进行检测,均取得了较好的检测结果,证明了基于迁移学习的深度学习在内镜图像多病灶检查任务中具有巨大潜力<sup>[9]</sup>。

目前自动的胶囊内镜病灶检测研究大多针对单一病灶,多病灶协同检测的研究还比较少,且存在的多病灶检测研究很少权衡模型的准确度与速度,难以适应临床中对多病灶协同检出的需求。为了自动识别胶囊内镜图像中的多种病灶,减少医务人员阅片的工作量,提高内镜诊断效率,本文结合深度学习网络与迁移学习技术对胶囊内镜图像中多种病灶的协同检测进行了深入研究,提出一种消化道常见病灶的协同检测方法。

## 1 多病灶协同分类检测网络

### 1.1 模型总体框架

卷积神经网络能够通过卷积核运算自动提取图像特征,为了获得多尺度的特征,网络结构经历了加深、加宽与加大模型输入尺寸的过程。但是,随着网络层数的加深、加宽,所带来梯度消失以及算法模型较为庞大引起过拟合等不利影响。为了便于胶囊内镜多病灶协同分类的嵌入式系统的实现,以及在算法准确率和计算时间达到一个较好的平衡,在此采用 EfficientNet 网络来平衡预测速度以及识别准确率。EfficientNet 是轻量级网络结构的代表,该网络利用残差模块增大了其网络深度,实现了更深层特征的提取,更加注重通道之间的关系,使得网络可以学习到不同通道特征的重要程度,关注有效信息,抑制无效信息,其高效的图像处理能力已经在 ImageNet 大型图像分类任务中得到证明,但目前还较少用于内镜图像的任务研究。模型训练与测试的流程如图 1 所示。

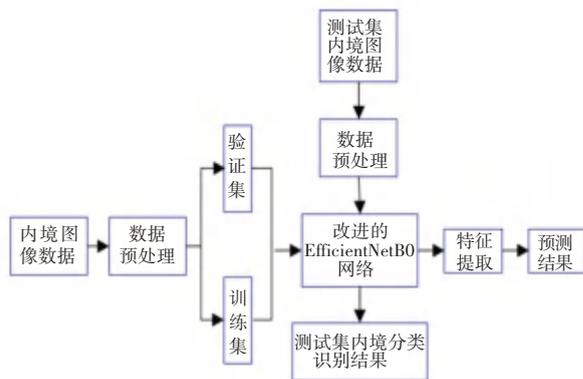


图 1 模型训练测试流程

Fig. 1 Model training and testing process

EfficientNet 网络的核心结构是图 2 所示的移动倒置瓶颈 (mobile inverted bottleneck convolution, MBConv) 模块和可分离卷积 (separable convolution, SepConv) 模块。这两种结构可以在兼顾网络参数量同时加强特征提取能力。

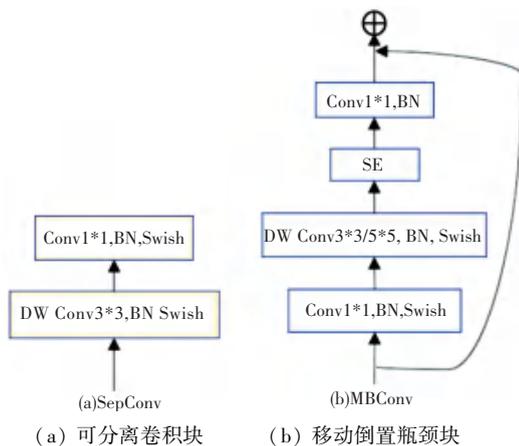


图 2 EfficientNet 主要模块结构图

Fig. 2 The main module structure of EfficientNet

可分离卷积块通常是将传统卷积操作拆分成深度卷积 (depthwise convolution, DW Conv) 和点卷积 (pointwise convolution, PW Conv) 两个步骤,是一种更高效的卷积方式;移动倒置瓶颈块主要由两个 Conv $1 \times 1$ 、DW Conv 和压缩激活模块 (Squeeze-and-Excitation block, SE block) 组成<sup>[10]</sup>,其中 DWConv 中卷积核大小可以为 3,也可以为 5,可以获取不同感受野大小的特征,此模块可以减少计算量,改变网络输出特征通道数,通过 SE block 关注输出通道特征信息,在部分模块中通过残差边强化网络输出特征重要性,提高网络特征提取能力。

### 1.2 多病灶分类网络

为了将 EfficientNet 网络模型更好地应用到胶囊内镜多病灶分类任务中,考虑到医学图像分类时数据集较小的现实,本文选取该网络结构体系的基线网络 EfficientNetB0,同时为了契合该胶囊内镜数据集,对其网络结构进行了调整,并在模型训练时对其微调。首先,为了降低模型计算量将裁剪后的图像统一调整为  $224 \times 224$  的大小后输入模型的输入层,输入卷积核为  $1 \times 1$  的卷积层进行特征粗提取,再经批量归一化层 (Batch Normalization, BN) 减少模型对初始化权重的依赖,并输入到可分离卷积块中,调整输出特征通道数;其次,16 个移动翻转瓶颈卷积网络块有序堆叠,依次是卷积核为  $3 \times 3$  的 MBConv3 重复 1 次,卷积核为  $5 \times 5$  的 MBConv5 重复 3 次,卷积核为  $3 \times 3$  的 MBConv3 重复 4 次,卷积核

为  $5 \times 5$  的 MBConv5 重复 7 次, 卷积核为  $3 \times 3$  的 MBConv3 重复 1 次; 最后, 通过可分离卷积块调整输出特征大小为  $(7, 7, 1\ 280)$ 。

此外, 在每个卷积层之后引入批量归一化层, 通过重新缩放对内镜图像特征图进行归一化, 防止梯度消失的同时快速收敛网络, 引入的残差结构通过内镜图像信息重利用, 有利于提高模型的稳定性, 加深网络层数, 提取到更深层次的特征。网络后端分别加以全局平均池化层, 将特征图转化为在某种意义上具有全局感受野的一维向量, 经具有 *ReLU* 激活函数的全连接层再次提取非线性特征, 最后经 Softmax 分类层输出, 通道数根据分类数量进行设置。为了实现分类血管畸形病灶图像、炎性病灶图像和正常消化道图像, 设置通道数为 3。表 1 给出了用于内镜图像病灶分类微调后 EfficientNet 网络的详细实现参数, 其中 MBConv3 代表卷积核为 3 的 MBConv 操作, MBConv5 代表卷积核为 5 的 MBConv 操作, 交替使用改变感受野获得不同大小的特征图。

表 1 微调后 EfficientNetB0 网络结构参数

Tab. 1 EfficientNetB0 network structure parameters after fine-tuning

Block name	Block number	Output size
Input	1	$224 \times 224 \times 3$
Conv	1	$112 \times 112 \times 32$
SepConv	1	$112 \times 112 \times 16$
MBConv3	1	$56 \times 56 \times 24$
MBConv5	3	$28 \times 28 \times 40$
MBConv3	4	$14 \times 14 \times 80$
MBConv5	4	$14 \times 14 \times 112$
MBConv5	3	$7 \times 7 \times 160$
MBConv3	1	$7 \times 7 \times 320$
SepConv	1	$7 \times 7 \times 1\ 280$
Average Pooling	1	1 280
Dense	1	256
Softmax	1	3

## 2 实验设计及结果分析

### 2.1 实验数据集

本文网络模型的训练和测试使用的是国际上公开的数据库 CAD-CAP, 该数据集来源于法国 12 家内镜检测中心收集的 4 174 个受检者的视频, 其中 1 480 个视频包含至少一个病灶, 2 694 个视频为正常消化道视频, 并由专业医师进行了图像诊断标注, 分为血管畸形病灶、出血点病灶、炎性病灶和正常 4

类。使用经过相同连续帧去冗余后发布在 Grand-challenge 网站上的 1 800 幅胃肠道图像, 包括 600 张血管畸形 (Vascular) 病灶图像、600 张炎性 (Inflammatory) 病灶图像和 600 张正常 (Normal) 消化道图像。

深度学习需要大量的样本用于训练以更新网络参数, 为了在小样本内镜数据集上使用深度学习的方法, 对内镜图像做裁剪与数据增强的预处理。内镜图像经过中心裁剪预处理归一化后进入网络输入端, 一方面可以为内镜消化道图像区域赋予强注意力, 减少了无关影像对实验结果的影响; 另一方面通过降低图像分辨率减少了模型训练的计算量。此外, 参考了机器视觉领域相似任务中的方法<sup>[11]</sup>, 结合内镜图像分类任务的自有特点, 模仿胶囊在体内不同拍摄角度与病灶形态与尺寸上的改变, 对图像进行旋转、平移、缩放、翻转等一系列操作, 同一张图像经过多种变换之后的结果如图 3 所示。

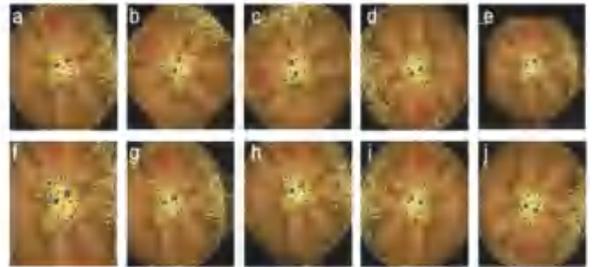


图 3 数据扩充

Fig. 3 Data augmentation

### 2.2 实验设计

本文基于 CAD-CAP 数据集研究内镜图像中的血管畸形病灶和炎性病灶的检出问题, 使用随机留出法来评估网络的分类性能。分别从 3 类数据中随机选出 20% 的数据构成测试集, 在余下的数据中随机选出 80% 的数据作为训练集, 20% 的数据作为验证集, 最终得到的训练集中包含 1 152 张图像, 验证集中包含 288 张图像, 测试集中包含 360 张图像。对训练集中的数据进行 9 倍数据增强, 验证集与测试集不使用数据增强。表 2 为扩充前后样本数据量。

表 2 网络模型训练及测试过程样本数量分布

Tab. 2 The distribution of the number of samples in the network model training and testing process

图像类型及 标签	原始图像			数据库
	训练验证集: 测试集 (0.8:0.2)			扩充后
	训练集	验证集	测试集	训练集
正常 0	384	96	120	3 456
血管畸形 1	384	96	120	3 456
炎性 2	384	96	120	3 456

训练和测试过程均在具有 CPU Intel® core™ i7-8700K @ 3.70 GHz 和 GPU NVIDIA GeForce GTX 1080Ti 的计算机上的实现,基于 Python 语言在 Keras 与 Tensorflow 深度学习框架,其中 Python 版本为 3.6.8, Tensorflow 版本为 1.14.0,keras 版本为 2.3.1。实验中网络最终使用 Softmax 分类层输出 3 类,计算公式(1)如下:

$$s_i = \frac{e^{\eta_i}}{\sum_{j=1}^c e^{\eta_j}} \quad (1)$$

其中,  $c$  为 3,表示分类数,  $\eta_i$  为第  $i$  个神经元的线性输出,经过 Sigmoid 激活函数后,输出预测值  $s_i$  被映射到  $(0,1)$ 。

为了评估本文所提监督方法对内镜图像的分类能力,采用准确率 (Accuracy)、精度 (Precision)、召回率 (Recall) 和 F1 分数 (F1 - score)、混淆矩阵作为性能评估标准。其中准确率、精度、召回率和 F1 分数指标的计算公式(2) ~ (5) 如下:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (4)$$

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} \times 100\% \quad (5)$$

其中,  $TP$  是真阳性样本数;  $FP$  是假阳性样本数;  $FN$  是假阴性样本数;  $TN$  是真阴性样本数。

同时使用混淆矩阵更加直观地观察测试结果。混淆矩阵可以用来可视化算法性能,其每一列代表预测值,每一行代表实际的类别,在预测分析时,预

测对的数据呈现在对角线上,从表格中可以清楚地观察分类结果。

### 2.3 实验结果分析

实验过程中模型的初始化参数均使用在 ImageNet 大型数据集上迁移学习的技术生成,在 CAD-CAP 测试集的 360 张图像上进行模型测试。使用的损失函数是交叉熵损失函数 (Cross Entropy Error Function),优化函数是 Adam,一阶动量设置为 0.9,二阶动量设置为 0.999,学习率设置为 0.000 1,批处理大小设置为 8,共训练 100 epochs。对比相同实验环境下 ResNet50、InceptionNetV4 网络的预测结果,验证轻量级网络对于多病灶协同检测任务的高效性与及时性。

实验结果见表 3,可以看出 3 个经过迁移学习微调的网络都取得了高于 90% 的准确度,达到了自动检测内镜图像中病灶的目的,其中改进后的 EfficientNetB0 对于正常图像的识别效果最好, F1-score 达到了 99.58%,而 InceptionV4 对于血管畸形和炎症图像的识别率较高,分别为 92.70% 和 93.17%。同时 InceptionV4 的整体准确率为最高的 95.00%,微调后的 EfficientB0 与之相当为 94.72%, ResNet50 炎性病灶的敏感度仅为 84.44%,难以平衡对多类病灶的检出敏感度。在测试时间消耗问题上,微调后的 EfficientB0 的消耗时间仅为 InceptionV4 的 1/4 (4.924/21.080),为 ResNet50 的 1/2 (4.924/8.929)。

综上所述,考虑到胶囊内镜一次采集便会获得较大数据量的特性,测试时间对于胶囊内镜多病灶分类识别也有着较大的意义,本文所提模型可以在分类准确率和测试时间达到更好的平衡,在各项指标上均体现出了一定优势。

表 3 CAD-CAP 数据集上网络模型的测试结果

Tab. 3 Test results of the network model on the CAD-CAP dataset

ResNet50	Precision/%	Recall/%	F1 score/%	Accuracy/%	Test time/s
Normal	98.36	100.00	99.17		
Vascular	95.15	81.67	87.89		
Inflammatory	84.44	95.00	89.41		
Total				92.22	8.929
InceptionV4	Precision/%	Recall/%	F1 score/%	Accuracy/%	Test time/s
Normal	100.00	98.33	99.16		
Vascular	95.58	90.00	92.70		
Inflammatory	89.92	96.67	93.17		
Total				95.00	21.080
EfficientNetB0	Precision/%	Recall/%	F1 score/%	Accuracy/%	Test time/s
Normal	100.00	99.17	99.58		
Vascular	94.74	90.00	92.31		
Inflammatory	89.76	95.00	92.31		
Total				94.72	4.924

模型测试结果的混淆矩阵如图 4 所示, 可以直观地看出微调后的 EfficientB0 没有将任何一类病灶图像预测为正常, 具有较高的异常查全率, 对于医学

检查而言, 这点尤为重要, 且对比另外两个网络的表现没有出现较多的预测失误。

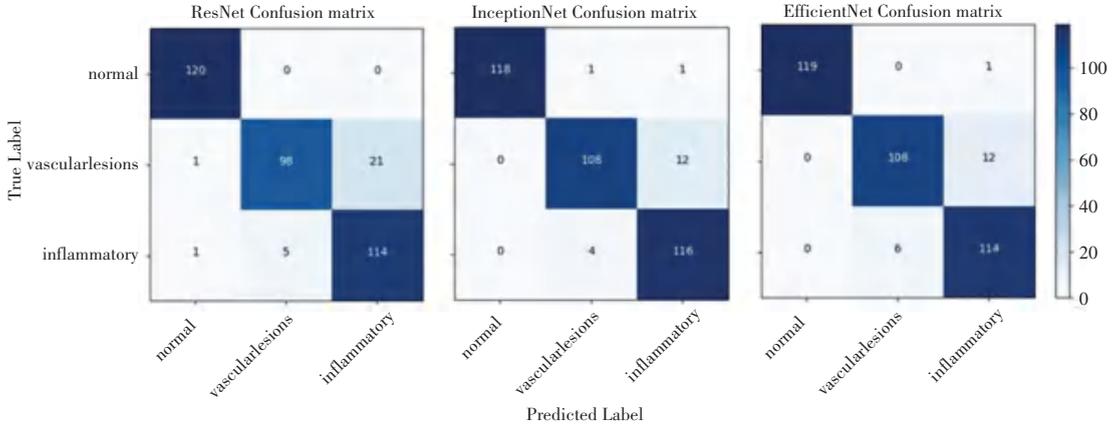


图 4 网络测试混淆矩阵结果的对比

Fig. 4 Comparison of the confusion matrix results of the network

### 3 结束语

本文针对胶囊内镜图像的多病灶协同检测识别准确率和识别时间这两大因素, 提出了一个基于轻量级网络迁移学习的方法, 利用轻量级网络 EfficientNetB0 在自然图像大数据集上学习到的特征知识, 通过载入预训练权重后对基础网络进行微调, 以更加适应该胶囊内镜样本数据集, 将该数据集输入至网络模型中, 快速输出内镜图像预测结果。对比其他网络, 所提方法具有较高的分类准确性和最快的检测效率, 模型准确度达到了 94.72%, 对正常图片的敏感度为 100%, 对血管畸形病灶的敏感度为 94.74%, 对炎症病灶的敏感度为 89.76%。预测 360 张胶囊内镜图像耗时 4.924 s, 检测速度为 73 帧/秒, 仅为 InceptionV4 的 1/4, 为 ResNet50 的 1/2, 以相同的速度检测 60 000 张内镜图像, 预计检测时间为 14 min, 证明了该模型在分类准确率和测试时间上达到了较好的效果, 也说明了轻量级网络在内镜病灶检测任务中存在着巨大潜力。

### 参考文献

- [1] SINGHI A D, KOAY E J, CHARI S T, et al. Early detection of pancreatic cancer: opportunities and challenges [J]. *Gastroenterology*, 2019, 156(7): 2024–2040.
- [2] WANG A, BANERJEE S, BARTH B A, et al. Wireless capsule endoscopy[J]. *Gastrointestinal endoscopy*, 2013, 78(6): 805–

815.

- [3] ESTEVA A, CHOU K, YEUNG S, et al. Deep learning-enabled medical computer vision[J]. *npj Digital Medicine*, 2021, 4(1): 1–9.
- [4] WEI J W, WEI J W, JACKSON C R, et al. Automated detection of celiac disease on duodenal biopsy slides: a deep learning approach[J]. *J Pathol Inform*. 2019, 10(1): 7–7.
- [5] XIA J, XIA T, PAN J, et al. Use of artificial intelligence for detection of gastric lesions by magnetically controlled capsule endoscopy[J]. *Gastrointestinal Endoscopy*, 2021, 93(1): 133–139.
- [6] OWAIS M, ARSALAN M, MAHMOOD T, et al. Automated Diagnosis of Various Gastrointestinal Lesions Using a Deep Learning-Based Classification and Retrieval Framework With a Large Endoscopic Database: Model Development and Validation [J]. *Journal of medical Internet research*, 2020, 22(11): e18563.
- [7] LEE J H, KIM Y J, KIM Y W, et al. Spotting malignancies from gastric endoscopic images using deep learning [J]. *Surgical Endoscopy*, 2019, 33(11): 3790–3797.
- [8] FAN S, XU L, FAN Y, et al. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images [J]. *Physics in Medicine and Biology*, 2018, 63(16): 165001.
- [9] OTANI K, NAKADA A, KUROSE Y, et al. Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network [J]. *Endoscopy*, 2020, 52(9): 786–791.
- [10] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-Excitation Networks[J]. *IEEE Transactions on Pattern Analysis And Machine Intelligence*, 2020, 42(8): 2011–2023.
- [11] SHORTEN C, KHOSHGOFTAAR T M. A survey on image data augmentation for deep learning[J]. *Journal of Big Data*, 2019, 6(1): 60.