

# Inception-ViT 模型的微型铣刀磨损状态预测研究

孙强, 李一全, 于占江, 李成超, 许金凯

长春理工大学跨尺度微纳制造教育部重点实验室

**摘要:** 微细加工过程中, 铣刀的磨损状态直接影响加工精度和被加工表面质量。为了预测刀具磨损状态, 提出一种 Inception-ViT 微型铣刀磨损预测模型。通过机器视觉检测装置获取铣刀后刀面磨损图像, 依据底刃直径减少量划分特征数据集, 通过数据增强方法扩充数据集。借助 Inception-ViT 模型对数据集进行训练, 建立微型铣刀磨损状态预测模型。为了验证模型具有优越性, 将本文提出的预测模型与常见分类模型对比分析。实验结果表明, Inception-ViT 模型在自建数据集上预测准确率最高可达 99.6875%。

**关键词:** 微细铣削; 磨损状态; Inception-ViT; 机器视觉

**中图分类号:** TG714; TH117.1; TP181

**文献标志码:** A **DOI:** 10.3969/j.issn.1000-7008.2022.01.001

## Prediction Research of Wear Status of Micro Milling Cutter by Inception-ViT Model

Sun Qiang, Li Yiquan, Yu Zhanjiang, Li Chengchao, Xu Jinkai

**Abstract:** In microfabrication, the wear status of the milling cutter directly affects the machining accuracy and the quality of the machined surface. An Inception-ViT milling tool wear prediction model is proposed in order to predict the tool wear status. A machine vision inspection device is used to obtain wear images of the rear face of the milling cutter and then a feature dataset is divided according to the amount of reduction in the bottom edge diameter, and the feature datasets are expanded by data enhancement methods. The datasets are trained by an Inception-ViT model to establish a prediction model of the wear status of micro milling cutters. In order to verify the superiority of the model, a comparative analysis of the prediction model and the common classification models is carried out. The experiment results show that the maximum accuracy of the Inception-ViT model reaches 99.6875% on the self-built datasets, and it is able to satisfy the requirements of actual tasks.

**Keywords:** microfabrication; wear status; Inception-ViT; machine vision

## 1 引言

微型铣刀的直径一般在 0.1 mm ~ 1 mm 之间, 加工过程中无法准确掌握铣刀磨损状态。如果保守估计刀具寿命, 铣刀无法被充分利用, 会降低生产效率。而过度估计会使铣刀严重磨损甚至破损, 造成巨大的经济损失和安全事故。因此掌握铣刀磨损状态、及时换刀或停机显得尤为重要<sup>[1]</sup>。

近年来随着深度学习理论的出现, 刀具磨损状态预测取得了突破性进展。刀具磨损状态预测方法主要是利用声发射、切削力、电流和振动信号来间接反映刀具磨损状态, 黄鹤翔等<sup>[2]</sup>提出基于 3-KMBS 的刀具磨损检测方法, 利用铣削加工中不同轴向的振动和声发射信号, 结合多隐层神经网络建立刀具磨损检测模型。Shi C. 等<sup>[3]</sup>提出多堆叠稀疏自动编码网络用于提取不同特征域的振动信号, 利用非线性回归函数对刀具磨损状态进行预测。何彦等<sup>[4]</sup>

提出基于长短时记忆卷积神经网络的刀具磨损在线监测模型, 通过提取振动、力和声发射信号, 并利用线性回归实现特征到磨损量化值的映射。陈启鹏等<sup>[5]</sup>提出基于深度门控循环单元神经网络的刀具磨损检测方法, 引入自注意力机制准确地预测刀具磨损状态。基于信号的刀具磨损状态预测方法可以实时动态监测磨损状态, 但是在提取信号特征过程中容易受环境因素的影响, 从而导致信号失真影响预测结果。针对该问题, 学者们提出利用机器视觉获取铣刀磨损图像, 再利用深度学习理论对刀具磨损图像分类来预测刀具磨损状态。Pagani L. 等<sup>[6]</sup>利用相机拍摄车削加工过程中产生的切屑, 再利用深度学习理论对切屑颜色进行分析, 间接反映刀具磨损状态。Bergs T. 等<sup>[7]</sup>利用卷积神经网络对铣刀磨损图像分类预测, 在测试集上经过测试, 模型的准确度达到 95.6%。

ViT 模型由 Dosovitskiy A. 等<sup>[8]</sup>在 ICLR2021 上提出, 模型在大型数据集的分类准确率高于 CNN 中的 EfficientNet 网络。本文参考 GoogleNet 中的 Inception 结构<sup>[9]</sup>, 提出 Inception-ViT 模型用于铣刀后刀面磨损状态预测分类, 通过与 ViT 模型对比分析,

**基金项目:** 国家重点研发计划项目(2018YFB1107403); 中国“111”计划项目(D17017); 吉林省科技发展计划项目(20190101005JH, 20180201057GX)

**收稿日期:** 2021 年 5 月

验证本文提出的预测模型在小规模自建数据集上预测分类的可行性。

## 2 模型设计

### 2.1 ViT 模型

ViT 模型中最重要的部分为 Transformer 编码器,主要由多头注意力机制层和多层感知机层两个子层组成,为了防止网络退化,每个子层都有残差连接<sup>[10]</sup>。ViT 模型的结构见图 1。

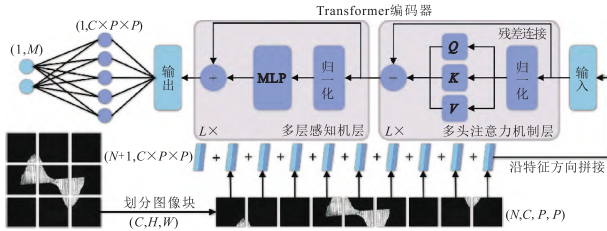


图 1 ViT 模型结构

ViT 模型对于图像分类任务的具体原理如下:

步骤①:首先将大小为  $(C, H, W)$  的图像分割为  $N$  个大小为  $(C, P, P)$  的不重叠图像块。其中,  $C$  为图像通道数;  $H$  和  $W$  为图像的高和宽;  $P$  为图像块的高和宽,参数关系式为  $N = HW/P^2$ 。

步骤②:将划分的图像块折叠为词嵌入向量,并根据具体位置分配位置嵌入向量。由于 Transformer 没有解码结构,需要添加一组随机向量用于分类。将所有的词嵌入向量和位置嵌入向量沿着特征方向拼接,得到  $N+1$  个大小为  $T$  的向量,  $T = CP^2$ 。

步骤③:多头注意力机制层由多个并行的自注意力机制组成,自注意力机制需要输入查询张量  $Q$ 、键张量  $K$  和值张量  $V$ 。三个张量的输入大小为  $(B, N+1, T)$ ,分别对应批次维度、特征数量和特征长度。如果自注意力机制的输出为  $Z$ ,那么转换过程可表示为

$$Z = \text{Softmax}\left(\frac{(QW_i^Q) \times (KW_i^K)^T}{\sqrt{d_i}}\right) \times VW_i^V \quad (1)$$

式中,  $W_i^Q$ ,  $W_i^K$  和  $W_i^V$  为  $Q$ ,  $K$  和  $V$  对应的权重矩阵;  $\sqrt{d_i}$  为特征维度相关系数。

将经过多头注意力机制后的输出和输入求和(残差连接),并按照式(2)~式(4)对输出结果进行层归一化处理,有

$$\mu = \frac{1}{C \times P \times P} \sum_{C,P,P} x \quad (2)$$

$$\sigma^2 = \frac{1}{C \times P \times P} \sum_{C,P,P} (x - \mu)^2 \quad (3)$$

$$y = \gamma \times \frac{x - \mu}{\sqrt{\sigma^2 + \varepsilon}} + \beta \quad (4)$$

式中,  $x$  和  $y$  分别为输入张量和输出张量;  $\varepsilon$  是一个小的浮点数常量(为了保证分母不为零);  $\gamma$  和  $\beta$  为缩放向量和平移向量,其元素的数目和输入张量通道数相等。

步骤④:将多头注意力机制层的输出作为多层感知器层的输入,经过层归一化和 MLP 后与输入相加(残差连接),输出大小为  $(B, N+1, T)$  的特征张量。对特征张量第二维度求均值,折叠后的输出张量大小为  $(B, T)$ 。

步骤⑤:经过  $L$  个 Transformer 编码器后通过多个全连接层后输出目标分类个数  $M$ ,最后通过 Softmax 输出每个类别的概率。

### 2.2 Inception-ViT 模型

Inception-ViT 模型主要参考了 GoogleNet 的 Inception 结构, Inception 结构见图 2。整个 Inception 模块从 Feature1 出发分成了 4 个分支,每个分支输出不同大小的特征张量。将 4 个分支的输出特征沿通道方向拼接在一起,输出 Feature2。由于 ViT 模型的主体部分均由全连接层组成,数据的第一维为批次大小,第二维为图像块的数量,图像块的通道、高度和宽度的维度折叠在一起形成第三维度。

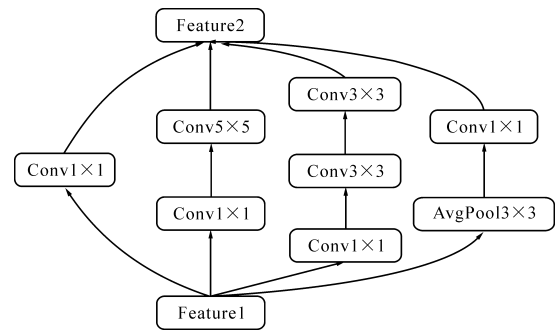


图 2 Inception 结构

对于 ViT 模型,图像块的大小会直接影响模型的预测准确率。如果图像块划分太小会遗失原始图像的大量特征信息,导致模型预测结果准确率较低。如果图像块划分太大会导致模型宽度变大,使得模型复杂度增加,进而导致模型训练速度较慢且容易发生拟合。因此本文提出一种 Inception-ViT 模型,通过融合多个 ViT 模型分支来提高模型的泛化能力,使用  $P_1$ ,  $P_2$  和  $P_3$  三种大小的图像块作为分支的 Inception-ViT 模型的原理见图 3。

对于输入大小为  $(C, H, W)$  的单张图像,按照  $P_1$ ,  $P_2$  和  $P_3$  划分不同图像块,图像块的个数分别为  $N_1$ ,  $N_2$  和  $N_3$ ,分别将其折叠成词嵌入向量和位置嵌入向量。通过 Transformer 编码器输出大小为  $Q_1$ ,  $Q_2$

和  $Q_3$  的特征向量,其中  $Q_i = P_i \times P_i, i = 1, 2, 3$ 。将 3 个分支的特征向量经过线性层输出大小相同为  $Q$  的特征向量,并通过变换转化为大小为  $(C, P, P)$  的特征张量。然后按照通道维度将 3 个分支进行拼接,得到输出特征张量,经过  $F1$ 、 $F2$ 、 $F3$  和  $F4$  后输出预测类型。为了防止预测模型过拟合,在线性层  $F1 \sim F4$  之间加入丢弃层和 ReLU 激活函数。

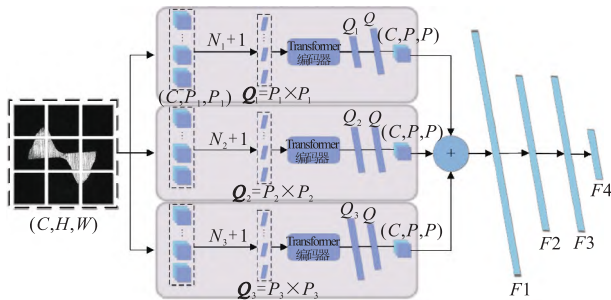


图3 Inception-ViT 模型原理

### 3 数据预处理

#### 3.1 数据集构建

预测模型的训练需要构建特征数据集,特征数据集主要由训练集和测试集组成。采用 1mm 的 TiN 涂层两刃硬质合金立铣刀铣削钛合金,并利用机器视觉检测装置拍摄微型铣刀后刀面图像,将后刀面图像作为特征数据集。特征数据集的数量为 6400,并按照 7 : 3 的比例划分为 4480 个训练集和 1920 个测试集。依据铣刀后刀面底刃直径减少量作为评价标准将特征数据集划分为无磨损、初期磨损、正常磨损和急剧磨损四种类型,分别用数字 0, 1, 2, 3 作为磨损状态类型标签。刀具磨损特征数据集见图 4。

#### 3.2 数据增强

为了提高预测模型的泛化能力,尽量考虑实际加工过程中微型铣刀后刀面图像的所有可能性,需要利用数据增强方法扩充特征数据集,常见的特征数据集增强方法有旋转变换、平移变换、加入噪声和比例变换等<sup>[11]</sup>。图像获取过程中刀具的旋转角度、位置是不确定因数,且受环境和设备因素的影响,获取到的图像可能会存在噪声。由于机器视觉检测装置的相机与光源参数已经固定,所以亮度、对比度对数据集影响较小,但是考虑到环境的复杂性,将对比度、亮度也纳入增强方法中。此外,通过比例变换可以模拟不同直径的微型铣刀的磨损图像,用于后续实验验证预测模型的泛化能力。本文采用的数据增强方法见图 5。

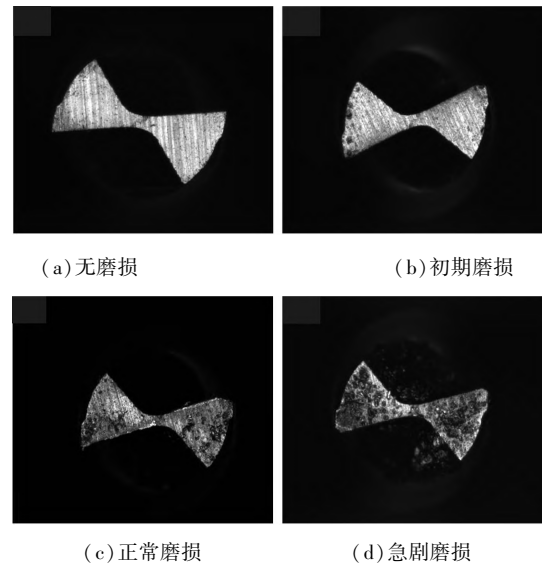


图4 特征数据集类别

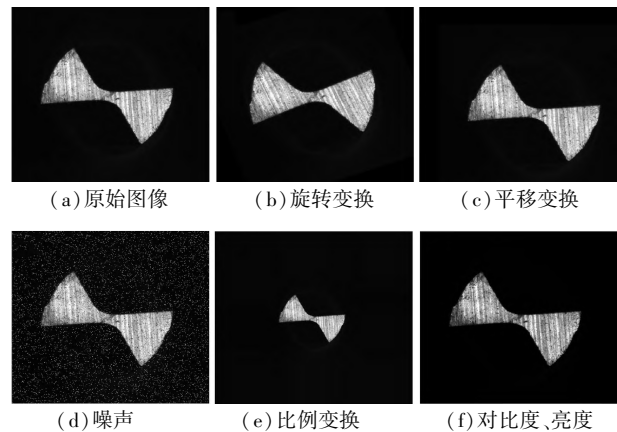


图5 特征数据集增强方法

## 4 实验与结果分析

### 4.1 实验环境与超参数

本文提出的预测模型和对比实验均在 Windows 操作系统上进行,使用 Pytorch 框架编写多融合 ViT 模型,具体实验环境参数见表 1。

表 1 实验环境配置

参数	操作系统	编程语言	框架	CUDA 版本
内容	Windows7	Python3.8	Pytorch1.7.1	10.2
参数	CPU 型号		GPU 型号	内存
内容	Intel Xeon CPU E5-2650 v4		Quadro M6000 24G	128GB

Inception-ViT 模型经多次实验后确定了最优参数,对模型的超参数设置如下:模型选择交叉熵损失函数、Adam 自适应优化器和固定步长衰减策略。训练迭代次数为 2240 次,迷你批次大小为 40,学习率为  $1 \times 10^{-4}$ ,固定步长衰减策略的步长为 10,衰减率为 0.1。

对于本文中的 Inception-ViT 模型,模型增加了



3 个分支结构,进一步扩充了模型宽度。但是模型太宽会增加模型复杂度,所以参照 ViT-base-16 将多融合 ViT 模型中每个分支的 Transformer 编码器层数设置为 4。对于输入为  $3 \times 224 \times 224$  的图像,可以划分出不同大小的图像块。本文依据  $8 \times 8, 14 \times 14, 16 \times 16, 28 \times 28$  和  $32 \times 32$  图像块共设计了 6 种模型,各模型参数见表 2。

表 2 不同分支结构模型参数表

模型	图像块大小			F1	F2	F3	F4
	分支 1	分支 2	分支 3				
模型 A	8	14	28	1764	1024	1024	4
模型 B	8	14	32	1764	1024	1024	4
模型 C	8	16	28	1764	1024	1024	4
模型 D	8	16	32	2304	1024	1024	4
模型 E	14	16	28	2304	1024	1024	4
模型 F	14	16	32	2304	1024	1024	4

4.2 评价指标

在机器学习领域,混淆矩阵常常用来评价监督学习中的模型性能。混淆矩阵的每一列代表预测类别,每一行代表数据的真实归属类别<sup>[12]</sup>。以二元分类问题为例,数据集存在肯定和否定两类记录,而分类模型对记录分类可能作出阳性或阴性判断。对于分类结果可作以下判别:作出正确判断的肯定记录  $TP$ 、作出错误判断的肯定记录  $FN$ 、作出正确判断的否定记录  $TN$  和作出错误判断的否定记录  $FP$ 。混淆矩阵结构见图 6。

真实值 \ 预测值	0	1
0	$TN$	$FN$
1	$FP$	$TP$

图 6 二元分类任务混淆矩阵

根据混淆矩阵可以计算准确率( $ACC$ )、精准度( $PPV$ )、召回率( $TPR$ )和  $F1$  值( $F1$ -score)等模型评价指标。针对本文自建特征数据集, $ACC$  用于衡量模型对 4 种类型刀具磨损的总体精度; $PPV$  和  $TPR$  用于衡量模型对于各磨损类型的预测精度;当  $PPV$  和  $TPR$  结果出现矛盾时,可以使用  $F1$  值来综合评价模型预测精度, $F1$  值越大说明模型预测精度越高<sup>[13]</sup>。各个评价指标的计算过程如下

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} \tag{5}$$

$$PPV = \frac{TP}{TP + FP} \tag{6}$$

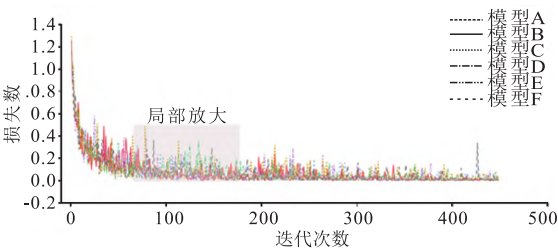
$$TPR = \frac{TP}{TP + FN} \tag{7}$$

$$F1 = \frac{2TP}{2TP + FP + FN} \tag{8}$$

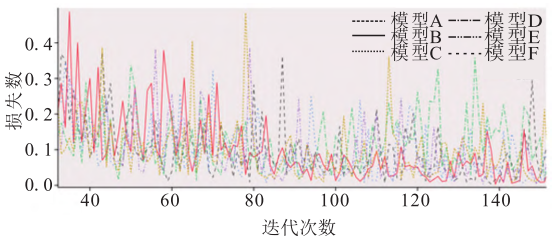
4.3 实验结果

(1)不同分支结构对比分析

对迭代次数间隔 5 采样并绘图,各分支结构模型训练过程中损失数和准确率随迭代次数变化曲线见图 7 和图 8。由图可见,针对本文自建数据集,模型 B 和 E 的收敛效果较好。

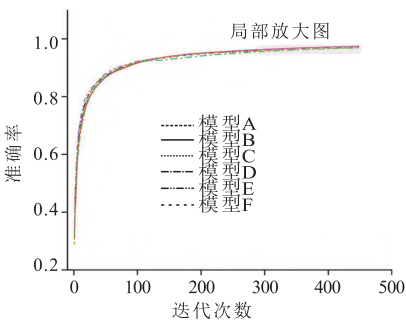


(a)原始图像

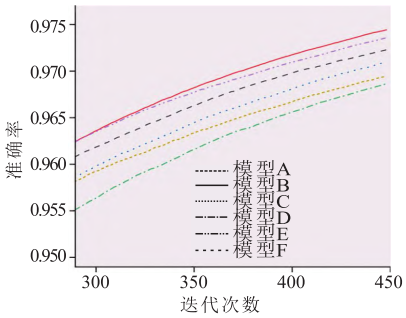


(b)局部放大图像

图 7 各模型训练过程损失数随迭代次数变化曲线



(a)原始图像



(b)局部放大图像

图 8 各模型训练过程准确率随迭代次数变化曲线

通过测试集数据对各个模型进行测试,测试各模型的准确率并预测单张图像的时间。每个模型分别进行 10 组实验,最后求出预测时间和准确率的平均值,其结果见表 3。从表中可以看出,模型 B 的测试准确率最高为 99.6875%,模型 A 的预测时间最少为 0.0173s。模型 E 的综合性能最好,在 0.0179s 的预测时间内可以达到 99.3229% 的预测准确率。

表 3 各模型测试结果

模型	模型 A	模型 B	模型 C
准确率 $ACC(\%)$	99.0104	99.6875	98.9583
预测时间 $t(s)$	0.0173	0.0200	0.0183
模型	模型 D	模型 E	模型 F
准确率 $ACC(\%)$	98.4375	99.3229	98.6458
预测时间 $t(s)$	0.0199	0.0179	0.0202

(2) Inception-ViT 与常见模型对比分析

为验证本文提出的 Inception-ViT 模型的优越性,在测试数据集上与 ViT、DeiT、TNT 和 ResNet - 50 等模型<sup>[14-17]</sup>进行对比分析,绘制混淆矩阵对比分析图(见图 9)。可见,在自建特征数据集上,本文提出的 Inception-ViT 模型测试准确率最高为 99.6875%,较其他模型提高 0.05% ~ 2.71%。且 Inception-ViT 模型对无磨损、初期磨损和正常磨损的预测较为理想,对于严重磨损状态仅有 6 组保守估计。

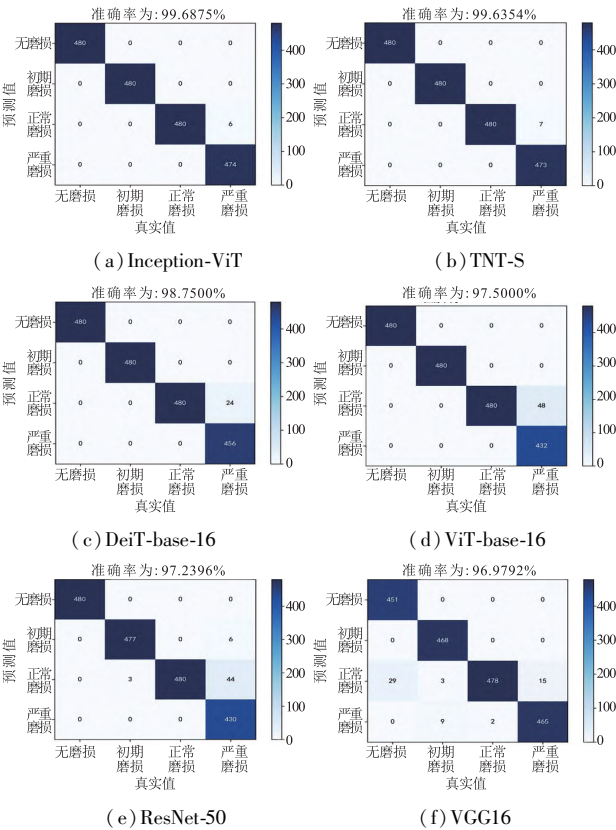


图 9 各模型测试结果对比分析

(3) 模型泛化能力测试结果

测试集上的测试结果因为数据集特征相似,无法准确评估模型的泛化能力。为了测试模型泛化能力,采用直径为 0.8mm 的微型铣刀(特征数据集所用铣刀直径为 1mm)进行模型泛化能力测试,总计 40 组数据。0.8mm 的铣刀磨损状态见图 10。

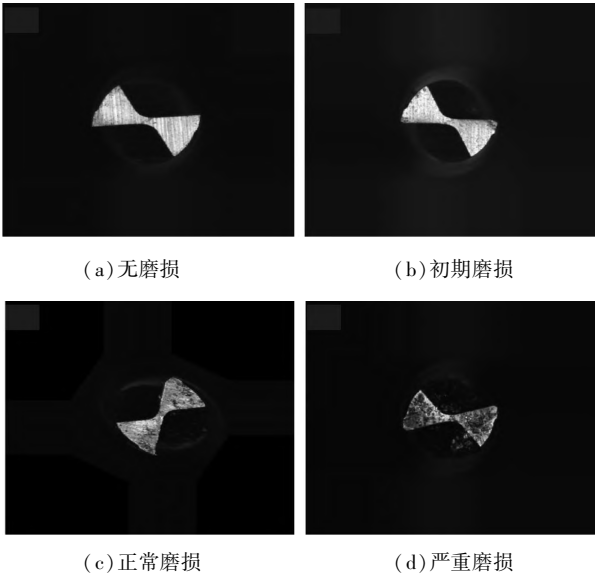


图 10 0.8mm 微型铣刀磨损状态

模型测试结果与人工显微镜检测对比结果见图 11。可见,模型的预测准确率为 92.5%,与在测试集上的误差小于 8%,具有一定泛化能力。

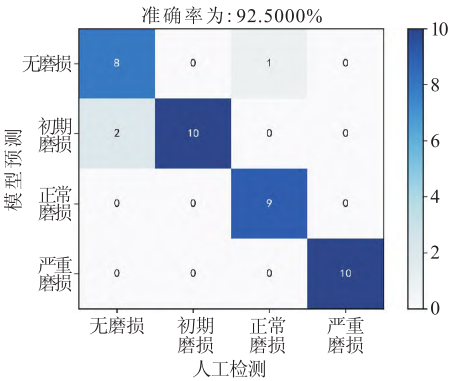


图 11 模型预测与人工检测结果对比

为了进一步验证预测模型的性能,根据图 11 所示混淆矩阵结果计算出微型铣刀各磨损状态下的精准度、召回率和  $F1$  值等性能评价指标,其结果见表 4。可见,预测模型对于正常磨损和严重磨损的预测精准度较高,且模型对于初期、正常和严重磨损状态的  $F1$  值大于 0.9,对于无磨损状态的预测效果较差。由于本文预测模型主要用于预测微型铣刀正常磨损和严重磨损状态,所以预测模型的性能较为理想。

表4 预测模型性能评价表

磨损状态	无磨损	初期磨损	正常磨损	严重磨损
精准度	0.8890	0.8330	1.0000	1.0000
召回率	0.8000	1.0000	0.9000	1.0000
F1 值	0.8421	0.9091	0.9474	1.0000

## 5 结语

针对微型铣刀磨损状态预测问题,依据 Inception 结构提出 Inception-ViT 模型。通过自建微型铣刀后刀面磨损图像特征数据集训练预测模型进行对比实验和泛化能力实验,验证模型的优越性和可行性,可以得出以下结论。

(1)设计了6种分支结构,通过对比实验可以得出模型B的准确率最高,模型A的预测时间最少,模型E的综合表现最好。

(2)通过与其他分类模型的对比实验验证了本文提出的 Inception-ViT 模型,测试准确率最高为99.6875%,较其他模型预测准确率最高提高2.71%。

(3)使用直径为0.8mm微型铣刀测试模型的泛化能力,预测准确率为92.5%,与测试集上的测试偏差小于8%,预测模型具有一定泛化能力。

## 参考文献

- [1]戴稳,张超勇,孟磊磊,等.采用深度学习的铣刀磨损状态预测模型[J].中国机械工程,2020,31(17):2071-2078.
- [2]黄鹤翔,张健,王鸿亮,等.基于3-KMBS的刀具磨损检测方法研究[J].组合机床与自动化加工技术,2020(10):90-93.
- [3]Shi C, Luo B, He S, et al. Tool wear prediction via multidimensional stacked sparse autoencoders with feature fusion[J]. IEEE Transactions on Industrial Informatics, 2020, 16(8):5150-5159.
- [4]何彦,凌俊杰,王禹林,等.基于长短时记忆卷积神经网络的刀具磨损在线监测模型[J].中国机械工程,2020,31(16):1959-1967.
- [5]陈启鹏,谢庆生,袁庆宽,等.基于深度门控循环单元神经网络的刀具磨损状态实时监测方法[J].计算机集成制造系统,2020,26(7):1782-1793.
- [6]Pagani L, Parenti P, Cataldo S, et al. Indirect cutting tool wear classification using deep learning and chip colour analysis[J]. The International Journal of Advanced Manufactur-

ing Technology, 2020, 111(3):1099-1114.

- [7]Bergs T, Holst C, Gupta P, et al. Digital image processing with deep learning for automated cutting tool wear detection[J]. Procedia Manufacturing, 2020, 48:947-958.
  - [8]Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth  $16 \times 16$  words: transformers for image recognition at scale[J]. ArXiv, 2020.
  - [9]林景栋,吴欣怡,柴毅,等.卷积神经网络结构优化综述[J].自动化学报,2020,46(1):24-37.
  - [10]Ashish Vaswani, Noam Shazr, Niki Parmar, et al. Attention is all you need[C]//I Guyon, U V Luxburg, S Bengio, et al. Advances in Neural Information Processing Systems 30. California: NIPS, 2017:5999-6009.
  - [11]张占军,彭艳兵,程光.基于CIFAR-10的图像分类模型优化[J].计算机应用与软件,2018,35(3):177-181.
  - [12]纪守领,李进锋,杜天宇,等.机器学习模型可解释性方法、应用与安全研究综述[J].计算机研究与发展,2019,56(10):2071-2096.
  - [13]Toensmeier P. Machine learning[J]. Plastics Engineering, 2018, 74(5):40-47.
  - [14]张建华,孔繁涛,吴建寨,等.基于改进VGG卷积神经网络的棉花病害识别模型[J].中国农业大学学报,2018,23(11):161-171.
  - [15]Çınar A, Yıldırım M, Eroğlu Y. Classification of pneumonia cell images using improved ResNet50 model[J]. Traitement du Signal, 2021, 38(1):165-173.
  - [16]Han K, Xiao A, Wu E, et al. Transformer in transformer[J]. ArXiv, 2021:2103.00112.
  - [17]Touvron H, Cord M, Douze M, et al. Training data-efficient image transformers & distillation through attention[J]. ArXiv, 2020:2012.12877.
- 第一作者:孙强,硕士研究生,长春理工大学跨尺度微纳制造教育部重点实验室,130022 长春市
- First Author: Sun Qiang, Postgraduate, Ministry of Education Key Laboratory for Cross-scale Micro and Nano Manufacturing, Changchun University of Science and Technology, Changchun 130022, China
- 通信作者:李一全,副教授,长春理工大学跨尺度微纳制造教育部重点实验室,130022 长春市
- Corresponding Author: Li Yiquan, Associate Professor, Ministry of Education Key Laboratory for Cross-scale Micro and Nano Manufacturing, Changchun University of Science and Technology, Changchun 130022, China