激光与光电子学进展 先进成像

基于轻量型卷积视觉Transformer的锑浮选工况识别

陈奕霏，蔡耀仪*，李诗文
湖南师范大学工程与设计学院，湖南 长沙 410083

摘要 依靠人工观测锑浮选泡沫特征进行锑浮选工况识别，主观性强、误差大，严重制约浮选性能。基于计算机视觉的识别方法成本低、效果好。针对以上问题，提出一种基于轻量型卷积视觉Transformer(L-CVT)的锑浮选工况识别方法。通过Transformer层的堆叠代替标准卷积中矩阵乘法来学习全局信息，将卷积中的局部建模更替为全局建模，同时引入轻量型神经网络MobileNetv2中的子模块，减少计算成本。所提方法解决了卷积神经网络(CNN)忽略浮选图像内部长距离依赖关系的问题，同时也弥补了视觉Transformer(VIT)缺乏归纳偏置的缺点。实验结果表明，基于所提方法的锑浮选工况识别准确率最高可达93.56%，明显高于VGG16、ResNet18、AlexNet等主流网络，为锑浮选数据在工况识别领域提供了重要参考。

关键词 机器视觉；锑浮选；工况识别；计算机视觉；轻量型卷积神经网络；视觉Transformer
中图分类号 TP391.4 **文献标志码** A **DOI**：10.3788/LOP213293

Working Condition Recognition Based on Lightweight Convolution Vision Transformer Network for Antimony Flotation Process

Chen Yifei, Cai Yaoyi*, Li Shiwen
College of Engineering and Design, Hunan Normal University, Changsha 410083, *Hunan, China*

Abstract It is highly subjective and has a large error to identify antimony flotation conditions by manually observing the characteristics of antimony flotation foam, which seriously restricts the flotation performance. The recognition method based on computer vision has low cost and good effect. In view of the above problems, a recognition method of antimony flotation conditions based on light-weight convolutional visual Transformer (L-CVT) is proposed. The stack of transformer layers replaces matrix multiplication in standard convolution to learn global information, replaces local modeling in convolution with global modeling, and introduces submodules in the lightweight neural network MobileNetv2 to reduce computational costs. The proposed method solves the problem that convolutional neural network (CNN) ignores the long-distance dependence within flotation images, and makes up for the lack of inductive bias of visual Transformer (VIT). The experimental results show that the accuracy of antimony flotation condition identification based on the proposed method can reach 93.56%, which is significantly higher than VGG16, ResNet18, AlexNet and other mainstream networks. It provides an important reference for antimony flotation data in the field of condition identification.

Key words machine vision; antimony floatation; condition recognition; computer vision; lightweight convolutional neural network; vision Transformer

1 引 言

矿产资源是人类社会赖以生存的重要物质基础。选矿是矿产资源加工中的一个重要环节，而泡沫浮选是一种应用广泛的选矿技术。浮选原理是利用矿物疏水性的差异将矿石和矿物分离，并加入浮选机来改变金属颗粒的疏水性，之后这些颗粒会附着到气泡上，最终形成聚集体。聚集体受浮力作用会上升到泥浆表面形成泡沫群[1]。浮选泡沫表观特征是评判浮选性能的重要依据，能直接反映浮选过程工况，工况的好坏将直接影响选矿回收率的高低。当前矿产资源匮乏，提高浮选性能、减少资源浪费、实现工况自动化生产至关重要。因此，基于视觉特征的浮选工况识别具有重要的研究意义。

浮选泡沫的表面视觉特征是浮选工况的直接指示

收稿日期：2021-12-20；**修回日期**：2022-01-07；**录用日期**：2022-01-17；**网络首发日期**：2022-01-27
基金项目：国家自然科学基金(61533021, 61134006, 61473319)、湖南省自然科学基金(2020JJ5366)
通信作者：*cyy@hunnu.edu.cn

器，但人工观察浮选泡沫特征进行工况识别，无法客观评价浮选状态，容易造成矿物原料大量流失、浮选药剂严重浪费等问题[2]。因此，大多浮选厂依靠提取泡沫表征来进行浮选的工况识别，比如浮选泡沫的颜色、大小、纹理等。Bartolacci等[3]利用多元图像分析的方法，提取泡沫RGB图像的颜色光谱变化特征来识别矿物品位；Ai等[4]提出一种通过气泡的大小和形状的联合分布对浮选工况进行判断的方法；阳春华等[5]提取浮选泡沫的纹理特征，再通过支持向量机对浮选工况进行分类。但这种将单一的泡沫视觉特征作为分类识别输入变量的方法，无法实现浮选工况的客观评价，导致识别准确率较低。为进一步提高浮选工况识别的准确率，梁秀满等[6]提出一种将泡沫亮点分布特征、图像灰度特征、Tamura纹理特征相融合的方法来表征浮选工况状态。虽然这种多特征融合的识别方法的确提升了浮选工况识别的准确率，但该方法的识别准确率受特征提取误差的影响较大，且忽视了特征间相互耦合的影响。因此，如何从浮选泡沫中提取丰富的特征信息，尚需进一步深入研究。文献[7]提出一种基于深度学习特征的浮选工况识别技术；Fu等[8]通过卷积神经网络(CNN)来提取浮选泡沫的特征，并在浮选图像识别任务中取得不错的效果；张进等[9]针对浮选加药状态检测困难等问题，提出一种基于多尺度CNN来提取浮选泡沫图像特征的方法。CNN通过卷积层来采集图像特征，具有一定的空间感知，但在实际浮选过程中，受现场恶劣环境、自然光照等因素干扰，浮选图像中泡沫大小不一、分布散乱。而CNN的空间感知是局部的，会忽略浮选图像内部的长距离依赖关系，并不能更全面、更准确地描述浮选泡沫表面的复杂特征信息。

为更全面地提取浮选泡沫的特征信息、提高浮选工况识别的准确率，本文提出一种基于轻量型卷积视觉Transformer(L-CVT)的锑浮选工况识别方法，引入缩放能力较强的视觉Transformer(VIT)[10]，通过Transformer[11]层的堆叠代替标准卷积中矩阵乘法(学习局部表示)的方式来学习全局表示。这种将Transformer视作卷积的方法，既能长距离捕获浮选泡沫特征，还能有效解决传统VIT在提取特征过程中会丢失图像内部像素空间顺序的缺点。结合轻量型神经网络MobileNetv2[12]中的子模块，利用模块内深度卷积的计算特性，大大减少因Transformer执行注意力操作带来的计算成本，解决实际工业中受设备资源的限制导致的锑浮选工况在线识别困难等问题。经实验比对，相较于VGG16[13]、ResNet18[14]、AlexNet[15]等网络而言，所提方法对锑浮选工况识别的准确率更高、计算成本更低。采用Grad-CAM[16]算法对4种网络进行特征图可视化，并解释说明了L-CVT网络提取特征的优点。

2　锑浮选流程及工况的类别

2.1　锑浮选流程

锑浮选流程如图1所示。将采集到的泥浆送入搅拌槽，通过搅拌槽中旋转磨机研磨成细粉状，再加入浮选药剂混合成浆料，最后将粗选矿库的尾矿送到锑浮选的工序中。完整的锑浮选流程包括粗选、清除和清洗。为了将锑矿物颗粒从无用材料中分离出来，操作人员在粗糙回路中注入药剂。在药剂作用下，粗糙回路中的气泡将矿物颗粒带到泥浆上层，形成泡沫层[17]。然后泡沫被运送到清洗槽，进一步提取和浓缩有价值的矿物。清洗槽排出的泥浆则被回收到更粗糙的回路中，泥浆下层流向辅助回收矿物的清除槽[18]，最终从清洗槽中获取锑精矿。

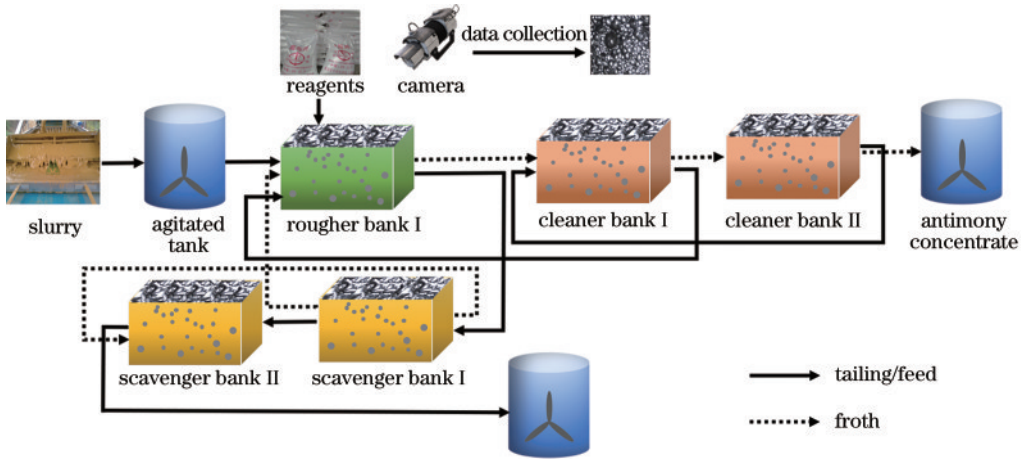

图1　锑浮选流程图

Fig. 1　Flow chart of antimony flotation

2.2　工况类别的确定

在图1中，粗选槽上方装有摄像机来采集锑浮选图像。经验丰富的操作人员通过化学分析来获取锑原矿中所含金属元素的百分比，即锑原矿品位。按锑原矿品位等级可以将锑浮选生产过程划分为异常、较差、合格、中、良、优等6种浮选工况，分别用类别Ⅰ、Ⅱ、Ⅲ、Ⅳ、Ⅴ、Ⅵ来表示。各类别代表性泡沫图像如图2所示。表1描述了这6种工况类别的具体信息。

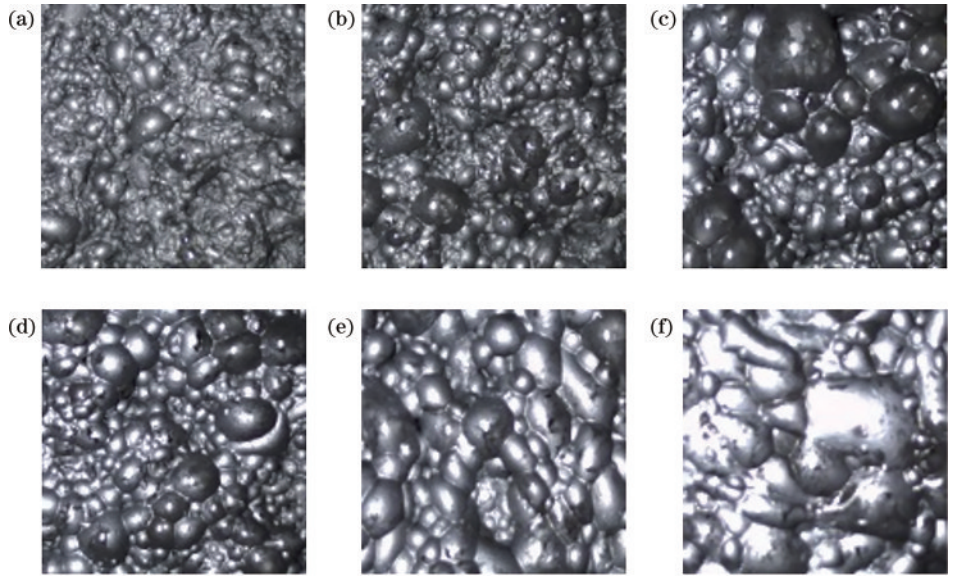

图 2　不同工况类别的浮选图像。(a) Ⅰ;(b) Ⅱ;(c) Ⅲ;(d) Ⅳ;(e) Ⅴ;(f) Ⅵ

Fig. 2　Flotation pictures of different working conditions. (a) Ⅰ; (b) Ⅱ; (c) Ⅲ; (d) Ⅳ; (e) Ⅴ; (f) Ⅵ

表 1　不同工况类别的特征描述

Table 1　Feature description of different operating conditions

Category	Flotation condition	Category feature description
Class Ⅰ	Abnormal	The bubbles are very sparse; the particle loading is much lower than normal and the bubbles are with gray appearance
Class Ⅱ	Poor	The bubbles are sparse; the particle loading is a little lower than normal and the bubbles are with gray-black appearance
Class Ⅲ	Qualified	The bubbles are medium in size and messy distributed; the particle loading is normal and the bubbles are with black appearance
Class Ⅳ	Medium	The bubbles are medium in size and evenly distributed; the particle loading is normal and the bubbles are with bright appearance
Class Ⅴ	Good	The bubbles are large in size and evenly distributed; the particle loading is higher than normal and the bubbles are with bright appearance
Class Ⅵ	Excellent	The bubble are the largest; the particle loading is much higher than normal and the froth is viscous; the bubbles are with water-shiny appearance

每种工况类别所对应的原矿品位如图 3 所示:第Ⅰ类锑原矿品位在 0.30%～0.92%;第Ⅱ类锑原矿品位在 0.92%～1.54%;第Ⅲ类锑原矿品位在 1.54%～2.16%;第Ⅳ类锑原矿品位在 2.16%～2.78%;第Ⅴ类锑原矿品位在 2.78%～3.40%;第Ⅵ类锑原矿品位在 3.40%～4.02%。

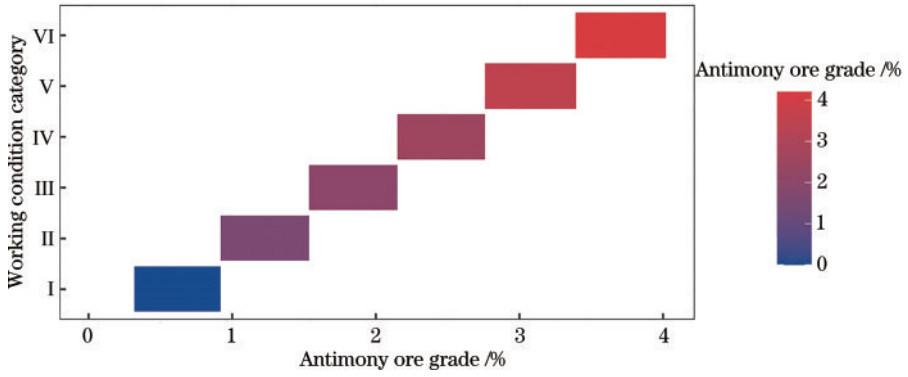

图 3　工况与品位之间的关系

Fig. 3　Relationship between working condition and grade

3　所提方法

3.1　基本网络的结构

L-CVT 由轻量卷积(L-Conv)模块、卷积视觉 Transformer(Conv-VIT)模块和多层感知机(MLP)构成。L-CVT 网络结构如图 4 所示。

将大小为 256×256 的锑浮选图像经过一个 3×3 的卷积核,然后输入由深度可分离卷积组成的 L-Conv 模块。其中,图 4(a)、(b)分别表示 L-Conv 模块中不同步长的深度卷积:当步长为 1 时,L-Conv 模块对特征图进行降维;当步长为 2 时,L-Conv 模块执行下采样来增大感受野。在连续经过 4 个 L-Conv 模块后,将特征图输入 Conv-VIT 模块,其模块结构如图 4(c)所示。Conv-VIT 模块通过卷积学习局部空间信息,通过 Transformer 层的堆叠来代替标准卷积中矩阵乘法,最后经过卷积层实现局部特征和全局特征的融合。为减少网络计算量和内存消耗,在经过多个 L-Conv 模块和 Conv-VIT 模块后,采用一个 1×1 的卷积核来压缩通道,最后通过 MLP 对锑浮选工况进行分类。L-CVT 网络既不改变每个编码浮选图像块及其内部像素的空间顺序,又能有效地将卷积中的局部建模更替为全局建模,使网络同时兼备 VIT 和 CNN 的优点。

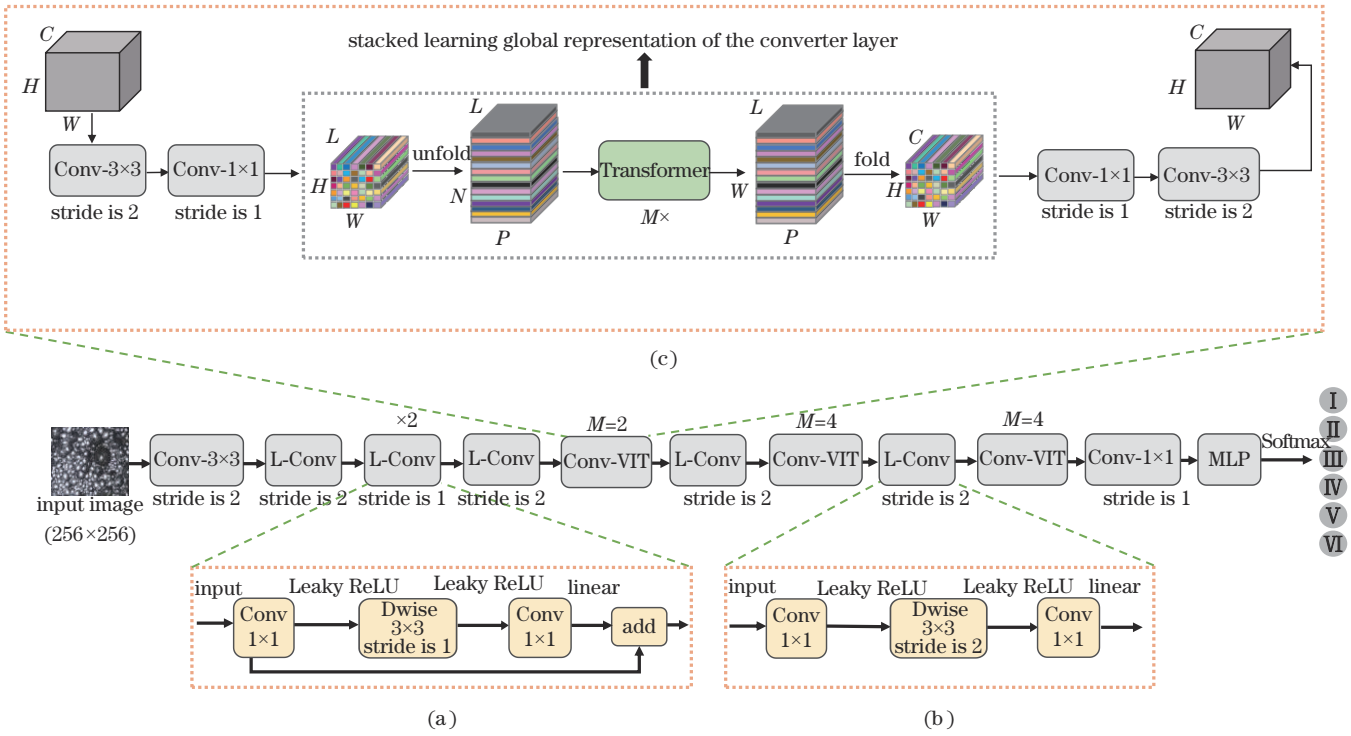

图 4　L-CVT 网络结构。(a)L-Conv 模块(深度卷积步长为 1);(b)L-Conv 模块(深度卷积步长为 2);(c)Conv-VIT 模块

Fig. 4　L-CVT network structure. (a) L-Conv module (depth convolution step size is 1); (b) L-Conv module (depth convolution step size is 2); (c) Conv-VIT module

3.2　L-Conv 模块

L-Conv 模块借鉴了 MobileNetv2 中的倒置残差结构。倒置残差结构的基本原理是先对浮选特征图进行通道扩张,再对浮选特征图进行通道压缩[19]。倒置残差结构的核心是深度可分离卷积。其中,深度可分离卷积包含一层深度卷积和一层逐点卷积,逐点卷积采用 1×1 的卷积来组合不同深度卷积的输出。将输入浮选图像与过滤器拆分为通道,并保证它们之间通道数目相同。对每个通道而言,将输入浮选图像与相对应的滤波器进行卷积,输出 2D 张量,并将其重叠在一起,最后通过逐点卷积来构造深度层输出的线性组合,完整操作如图 5 所示。值得注意的是,由于深度卷积本身的计算特性决定其不能改变通道数目,为避免上一层输出通道数目较少时,深度卷积只能在低维空间提取特征的缺点,在深度卷积前增加一个逐点卷积来增加通道数目,如图 4(a)所示。

本实验对倒置残差结构进行一定的改进。使用 Leaky ReLU 激活函数代替原来结构中的 ReLU6 激活函数。因为 ReLU6 的输入值为负时,输出始终为 0,其一阶导数也始终为 0,这会导致神经元不能更新参数,而 Leaky ReLU 激活函数保留了部分负轴的值,使得负轴的信息不会全部丢失。Conv 模块对两个子层均使用 Leaky ReLU 激活函数和 BatchNorm[20]归一化。每个输入通道均有一个滤波器的深度卷积,其计算方式为

$$M_{k,l,m}=\sum_{i,j}\hat{K}_{i,j,m}F_{k+i-1,l+j-1,m},\tag{1}$$

式中:$\hat{K}$ 为深度卷积核;F 为输入特征映射;M 为输出特征映射。整个式子表示在深度卷积核 $\hat{K}$ 中,第 m 个

滤波器应用在输入特征映射 F 中第 m 个通道，得到滤波后的输出特征映射 M。

给定深度卷积核大小为 $D_k \times D_k$，输入通道数为 C_i，输入特征映射大小为 $D_f \times D_f$，输出通道数为 C_o，那么深度可分离卷积的计算量可表示为

$$L_{DW} = D_f D_f D_k D_k C_i , \tag{2}$$

$$L_{PS} = D_f D_f C_o C_i , \tag{3}$$

$$L_{DWPS} = L_{DW} + L_{PS} , \tag{4}$$

式中：L_{DWPS}、L_{DW}、L_{PS} 分别代表深度可分离卷积的计算量、深度卷积的计算量、逐点卷积的计算量。为对比深度可分离卷积与标准卷积的计算成本，给定标准卷积核大小为 $D_k \times D_k$，输入通道数为 C_i，输入特征映射大小为 $D_f \times D_f$，输出通道数为 C_o，标准卷积的计算量为

$$\begin{cases} L_{Std} = D_k D_k D_f D_f C_i C_o \\ n = \dfrac{L_{DWPS}}{L_{Std}} = \dfrac{1}{C_o} + \dfrac{1}{D_k^2} \end{cases} , \tag{5}$$

式中：L_{Std} 表示标准卷积的计算量；n 表示倍数。通常情况下输出通道数 C_o 较大，所以深度可分离卷积相较于标准卷积而言，能够降低约 $\dfrac{1}{D_k^2}$ 的计算量。

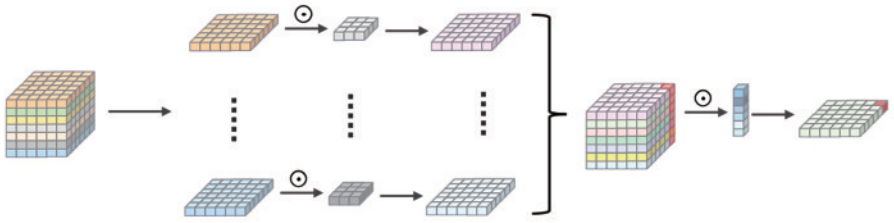

图 5　深度可分离卷积

Fig. 5　Depth separable convolution

3.3　Transformer

Transformer 是 Conv-VIT 模块的核心，它包含两个子层，多头自注意力层和前馈连接层，每个子层后面都加上残差连接和正则化层，其内部结构如图 6(a)所示。多头注意力机制是 Transformer 的重要组成部分，结构如图 6(b)所示，它由多个自注意力相连接，为注意力提供多种可能性。注意力机制的本质在于使计算机模仿人类观察事物的方式[21]。每个自注意力将需要编码的特征向量与 3 个训练过程中学习得到的权值矩阵 $\boldsymbol{W}^Q$、$\boldsymbol{W}^K$、$\boldsymbol{W}^V$ 相乘，获得 3 个新的向量 $\boldsymbol{Q}$(query)、$\boldsymbol{K}$(key)、$\boldsymbol{V}$(value)。本实验使用缩放点积注意力来计算相似度：

$$\text{Attention}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d}}\right)\boldsymbol{V}, \tag{6}$$

式中：d 表示 $\boldsymbol{Q}$ 和 $\boldsymbol{K}$ 的维度，除以 $\sqrt{d}$ 可以防止梯度消失，并用 Softmax 函数来获取 $\boldsymbol{V}$ 的权重。最后将多个单头注意力机制连接起来得到多头注意力机制，其计算方式为

$$\begin{cases} \text{multi-head}(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = \text{Concat}(\text{head}_1, \cdots, \text{head}_h)\boldsymbol{W}^O \\ \text{head}_i = \text{Attention}(\boldsymbol{Q}\boldsymbol{W}_i^Q, \boldsymbol{K}\boldsymbol{W}_i^K, \boldsymbol{V}\boldsymbol{W}_i^V) \end{cases} , \tag{7}$$

式中：$\boldsymbol{W}^O$、$\boldsymbol{W}_i^Q$、$\boldsymbol{W}_i^K$、$\boldsymbol{W}_i^V$ 代表线性变化时的参数矩阵。多头注意力机制的本质是将 $\boldsymbol{Q}$、$\boldsymbol{K}$、$\boldsymbol{V}$ 这 3 个参数进行多次拆分，每组拆分参数映射到高维空间的不同子空间中计算注意力权重。经多次并行计算，最终合并所有子空间中的注意力信息。

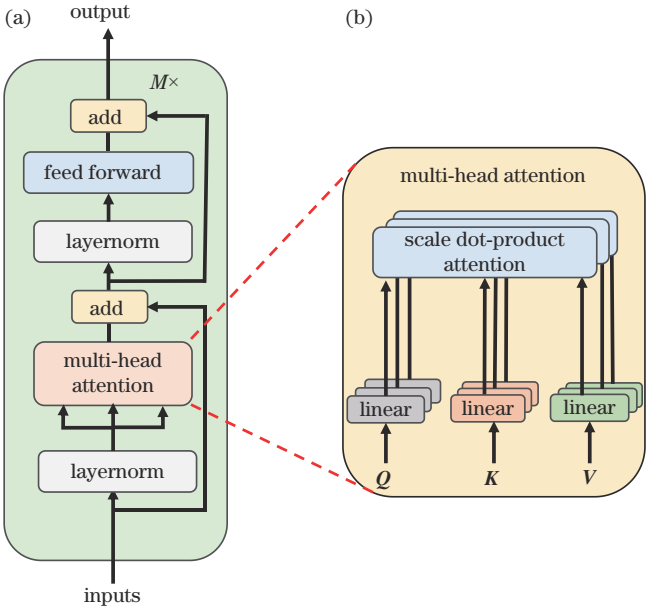

图 6　Transformer。(a)Transformer 结构；(b)多头注意力机制

Fig. 6　Transformer. (a) Transformer structure; (b) multi-head attention

3.4　卷积视觉转换器模块

Conv-VIT 模块将 Transformer 当作卷积来学习全局表示，如图 4(c)所示。标准卷积通常由展开、矩阵乘法和折叠这 3 个连续操作构成。Conv-VIT 模块与卷积相似，通过 Transformer 层的堆叠来代替卷积中的矩阵乘法，使整模块由局部处理转为更深的全局处理。

给定输入浮选特征图 $\boldsymbol{X} \in \mathbb{R}^{H \times W \times C}$，$H$ 和 W 分别表示特征图的高度和宽度，C 为通道数。首先使用 3×3 的卷积和 1×1 的卷积核对输入特征图进行卷积，得到 $\boldsymbol{X}_i \in \mathbb{R}^{H \times W \times L}$，其中，3×3 的卷积用于学习局部空间信息，1×1 的卷积将输入特征投影到高维空间。为获得浮选图像内部长距离依赖关系，采用具有多头自注意

的VIT先将$\boldsymbol{X}_i\in\mathbb{R}^{H\times W\times L}$平展为$N$个不重叠的图像块$\boldsymbol{Q}\in\mathbb{R}^{P\times N\times C}$。其中，$P=w\cdot h$，图像块的数目$N=\frac{W\cdot H}{P}$，$w$和$h$分别指浮选图像块的宽和高。给定浮选图像块中每个像素$z\in\{1,2,\cdots,P\}$，则通过Transformer进行编码的过程可表示为

$$\boldsymbol{Y}(z)=T_{\text{Transformer}}\left[\boldsymbol{Q}(z)\right],\ 1\leqslant z\leqslant P,\tag{8}$$

式中：$\boldsymbol{Y}(z)$表示每个像素经过Transformer编码后的输出结果。本实验网络与传统VIT网络不同，并不会丢失浮选图像块及其内部像素的空间顺序。将输出$\boldsymbol{Y}\in\mathbb{R}^{P\times N\times L}$折叠得到$\boldsymbol{Y}_j\in\mathbb{R}^{P\times N\times L}$后，再通过1×1的卷积和3×3的卷积将$\boldsymbol{Y}_j$投影到低维空间，融合局部特征和全局特征。$\boldsymbol{Q}(z)$使用卷积对3×3区域局部信息进行编码，但$\boldsymbol{Y}(z)$对$P$个图像块中第$z$个位置的全局信息进行编码，所以能够对$\boldsymbol{X}_i\in\mathbb{R}^{H\times W\times C}$中的全局信息进行感知，如图7所示。

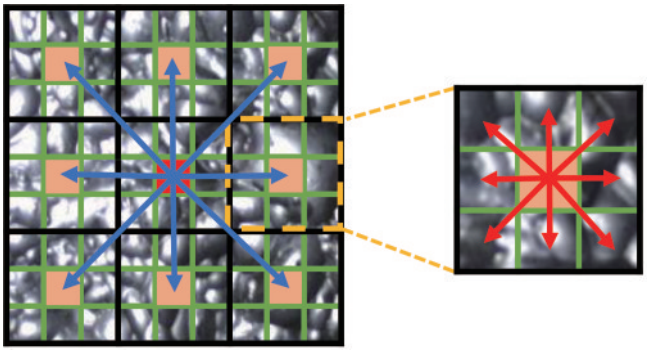

图7　Conv-VIT模块对像素信息的全局表示

Fig. 7　Global representation of pixel information by Conv-VIT module

在图7中，黑色网格和绿色网格中的每个单元分别表示一个锑浮选图像块和一个像素。锑浮选图像中心的红色像素通过转换器编码，处理其余图像块中相应像素的位置（用橙色表示）。橙色像素在使用卷积后，对自身周围像素的信息进行编码，从而使得这些像素能够对锑浮选图像中所有像素的信息进行编码。

3.5　MLP

MLP的层与层之间是全连接的，最底层是输入层，中间是隐藏层，最后是输出层。本实验采用PReLU非线性激活函数来增加网络的非线性拟合能力，其计算方式为

$$\mathrm{PReLU}(x_i)=\begin{cases}x_i, & x_i>0\\ a_i x_i, & x_i\leqslant 0\end{cases},\tag{9}$$

式中：x_i表示非线性激活函数在第i个通道的输入；a_i是权重系数。将输出特征输入MLP，利用隐藏层建立输入层和输出层之间的连接，对输入特征进行非线性变换，输出层的神经元数量为工况类别数量。通过Softmax函数，将分类值转换为概率分布，取最高概率标签为最佳结果，最终实现锑浮选的工况识别：

$$S_j=\frac{\exp(x_j)}{\sum_{i=1}^{n}\exp(x_i)},\quad j=1,2,\cdots,n,\tag{10}$$

式中：x_i是指浮选图像中第i个标号，浮选图像中总共有个n标号；S_j是x_j的Softmax函数的输出结果。

3.6　网络优化

为加快网络的收敛速度，防止出现过拟合，本实验在全连接层中加入Dropout[22]的优化方法，这是一种在神经网络的隐藏单元中加入噪声来调整网络的方式。丢弃隐藏层中一些节点的连接，采用部分连接的方式，让部分隐藏层的节点不再工作，以此避免某些特征只能在固定组合中才能生效的缺点，让模型学习一些普遍的共性。优化器选择Adam算法，这是一种学习率自适应的优化算法，其显著优点在于实现简单、计算高效，超参数具有很好的解释性。

4　实验结果和分析

4.1　数据采集

为验证所提方法，以湖南邵阳某选矿工厂作为研究对象，数据采集现场如图8所示。采集浮选图像所采用的设备是AVT工业相机和Kowa镜头。相机为G-223B/C的AVT-Manta系列工业相机，其分辨率大小为2048×1088，帧速为53.7 frame/s，内存大小128 MB。镜头为LM35HC的Kowa-HC系列镜头，焦距为35 mm。实验共采集6种不同工况类别的锑浮选图像。原始数据集总共包含9000张浮选图片，其中，每类工况的训练集各含1200张，测试集各含150张，验证集各含150张。

但浮选现场环境恶劣、光照强度不均，导致部分采

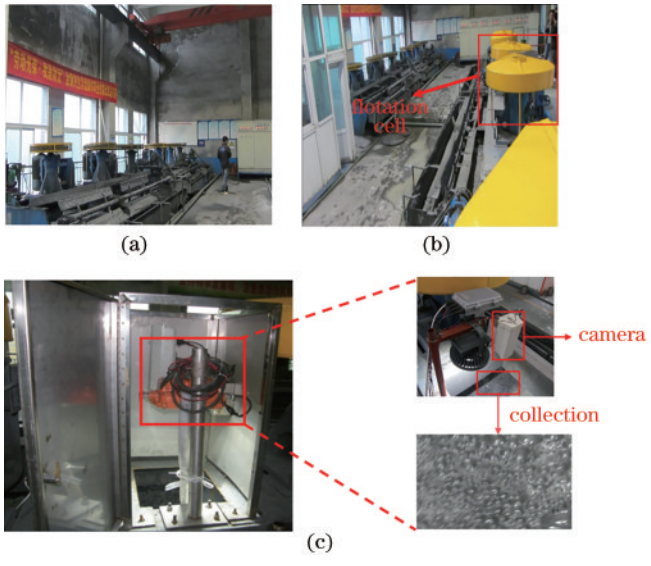

图8　浮选数据采集过程。(a)浮选现场；(b)浮选槽；(c)浮选数据的采集终端

Fig. 8　Process of flotation data collection. (a) Flotation site; (b) flotation tank; (c) collection terminal of flotation data

集的浮选图像存在一定的噪声，而噪声点对浮选图像造成干扰，使得图像变得不清晰。因此，采用高斯平滑滤波器[23]对浮选图像进行预处理，一定程度上可以降低恶劣环境对图像的影响。

4.2　数据增强

实验采集样本相对较少，仍存在过拟合的风险。针对以上问题，借助数据增强来减轻过拟合程度。图像增强基于现有的训练数据生成随机图像，能很好地提升训练模型的泛化能力。本实验采用的数据增强方法分别是图像翻转、MixUp[24]、CutMix[25]，并将增强处理后的浮选图像作为训练集数据。

1）图像翻转变化，随机将部分训练集的浮选泡沫图像进行水平翻转、垂直翻转、顺时旋转 90°，效果如图 9 所示。

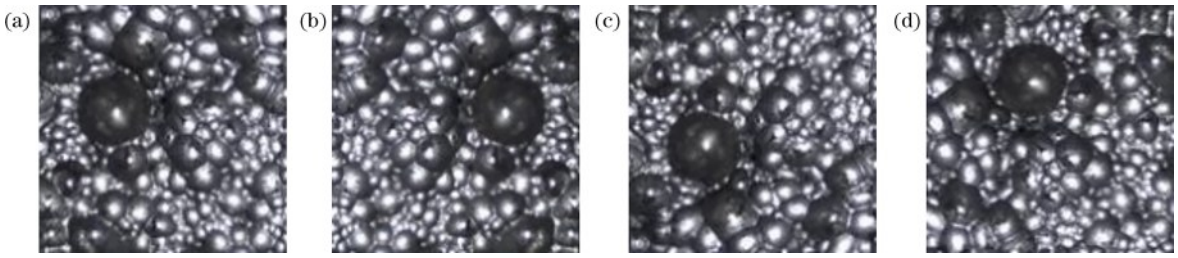

图 9　图像翻转变换。(a)原图；(b)水平翻转；(c)垂直翻转；(d)顺时旋转 90°

Fig. 9　Image flip transformation. (a) Original image; (b) horizontal flip; (c) filp vertically; (d) rotate clockwise 90°

2）MixUp 是一种对图像进行混合的增强算法，能够将不同工况类别间的浮选泡沫图像进行随机混合，在训练过程中不会出现非信息像素，从而提高训练效率。假设给定 B_{X1} 是一个 batch 样本，该样本所对应的标签为 B_{Y1}；B_{X2} 是另外一个 batch 样本，该样本所对应的标签为 B_{Y2}，通过参数 α 和 β 的 Beta 分布计算出混合系数 λ。MixUp 算法的原理可以描述为

$$\lambda = B_{\mathrm{Beta}}(\alpha,\beta), \tag{11}$$

$$B_{\mathrm{Mixup\_Batch}_X} = \lambda \times B_{\mathrm{Batch}_{X1}} + (1-\lambda) \times B_{\mathrm{Batch}_{X2}}, \tag{12}$$

$$B_{\mathrm{Mixup\_Batch}_Y} = \lambda \times B_{\mathrm{Batch}_{Y1}} + (1-\lambda) \times B_{\mathrm{Batch}_{Y2}}, \tag{13}$$

式中：$B_{\mathrm{Mixup\_Batch}_X}$、$B_{\mathrm{Mixup\_Batch}_Y}$ 分别代表混合后的 Batch 样本和该混合样本所对应的标签。在实验中，Beta 分布的参数 α 和 β 设置为 0.5，增强效果如图 10 所示。

3）CutMix 将浮选泡沫图像的一部分区域剪裁掉，随机填充训练集中其他浮选泡沫图像的部分区域像素值，分类结果按一定比例分配。整个过程不会出现图像混合后不自然的情况，能有效提升模型分类的表现。假设 X_1 和 X_2 是来自两个不同训练集的样本，Y_1 和 Y_2 是分别对应的标签值，经过 CutMix 算法后得到新的训练集样本 X^* 和对应标签值 Y^*。CutMix 算法的原理可以描述为

$$X^* = NX_1 + (1-N)X_2, \tag{14}$$

$$Y^* = \lambda Y_1 + (1-\lambda)Y_2, \tag{15}$$

式中：为了剪裁掉部分区域和进行填补的二进制掩码 $N \in \{0,1\}^{W \times H}$；$\lambda$ 与 MixUp 算法式(11)一样，都属于 Beta 分布。为了对二进制掩码进行采样，需要对剪裁部分的边界框 $L=(Z_x, Z_y, Z_w, Z_h)$ 进行采样，裁剪部分的边界框采样公式为

$$\begin{cases} Z_x \sim \mathrm{Unif}(0, W) \\ Z_w = W\sqrt{1-\lambda} \end{cases}, \tag{16}$$

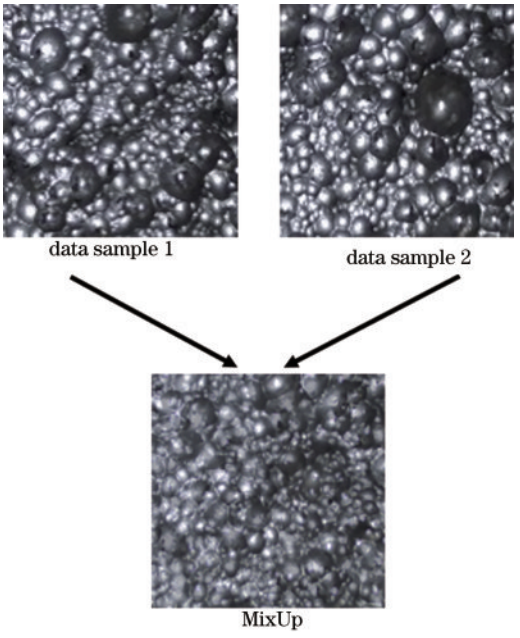

图 10　MixUp 效果图

Fig. 10　MixUp rendering

$$\begin{cases} Z_y \sim \mathrm{Unif}(0, H) \\ Z_h = H\sqrt{1-\lambda} \end{cases}。 \tag{17}$$

保证剪裁区域的比例 $\frac{Z_w Z_h}{WH} = 1-\lambda$，确定好剪裁区域 L 后，将制掩码 N 中的剪裁区域 L 置 0，其他区域置 1。然后将样本 1 中剪裁区域移除，将样本 2 中的剪裁区域剪裁后填入样本 1 中。在实验中，剪裁区域设置为 1∶1。增强效果如图 11 所示，图中方框表示裁剪区域 L。

为进一步探究数据增强的有效性，在相同的训练环境下，结合这 3 种数据增强方法，并进行对比，实验结果如表 2 所示。

从表 2 中前 3 组实验结果可以看出：CutMix 方法

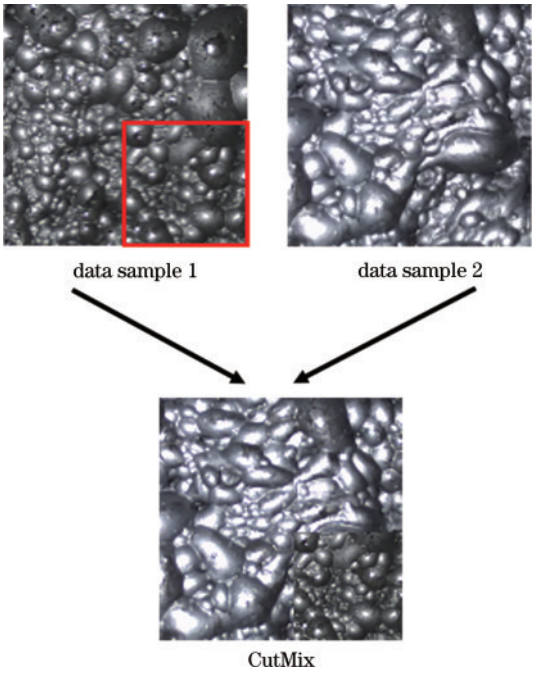

图 11　CutMix效果图

Fig. 11　CutMix rendering

表 2　数据增强对比试验

Table 2　Experiment of data augmentation comparison

Method	Top-1 accuacry /%
Flip	89.64
MixUp	90.83
CutMix	91.39
Filp+MixUp+CutMix	93.56
None	86.55

可以较大程度地提升网络的识别准确率；第 4 组实验则说明，综合多种增强方法，能进一步提升网络性能。实验结果表明，当样本数较少时，将数据增强作为数据扩充的方法，能在一定程度上减少过拟合的倾向，并有效提升网络的泛化性和鲁棒性。

4.3　实验配置

本实验使用的编程语言为 Python 3.8，采用框架是 PyTorch 1.7.1，实验运行环境为 Intel(R)Core(TM) i5-10210U CPU(1.60 GHz)、Nvidia RTX2080Ti GPU 和 Intel(R)Xeon(R)E5-2640 CPU(2.40 GHz)。为确保实验的公平性，所有对比模型均使用相同的实验环境及配置。将输入的浮选图像大小设置为 256×256。为加快网络的训练速度，减轻计算量，采用归一化处理，将数据归一化到[−1,1]。

实验采用总体准确率作为评价指标，批量大小设置为 64，实验迭代轮数(epoch)设置为 50。学习率是一个非常重要的超参数，它表示网络权重更新的速率：设置过大容易造成代价函数的波动，导致实验结果不够准确；设置过小则网络的收敛效果欠佳，导致训练时间过长[26]。因此，本实验等间隔调整学习率，将初始学习率设置为 1.0×10^{-4}，设置学习衰减率为 0.1，间隔步长为 10。最优控制器选择 Adam 优化器，损失函数选择交叉熵损失函。丢弃法中折损率设置为 0.5。网络参数如表 3 所示，其中，FLOPs 表示浮点运算数量，Params 表示参数量。

表 3　L-CVT 网络参数

Table 3　Network parameters of L-CVT

Layer name	Output size	Output channels	Number
Conv-3×3	128×128	32	1
L-Conv(stride is 2)	64×64	48	1
L-Conv(stride is 1)	64×64	48	2
L-Conv(stride is 2)	32×32	64	1
Conv-VIT(M=2)	32×32	64	1
L-Conv(stride is 2)	16×16	96	1
Conv-VIT(M=4)	16×16	96	1
L-Conv(stride is 2)	8×8	128	1
Conv-VIT(M=4)	8×8	128	1
Conv-1×1	8×8	384	1
MLP	1×1	6	1
FLOPs: 6.01×10^{10}	Params: 2.33 MB		

4.4　实验结果及网络性能评估

在相同实验环境及配置下，对锑浮选数据进行工况识别。采用 AlexNet、VGG16、ResNet18 这 3 种识别网络与所提网络进行对比实验。表 4 为这 3 种对比网络的主要参数。

表 5 列出了基于 L-CVT、AlexNet、VGG16、ResNet18 这 4 种不同模型的锑浮选工况识别准确率。其中，基于 L-CVT 的工况识别准确率为 93.56%，在 4 种网络中最高。L-CVT 相较其他 3 种网络而言，其识别准确率分别提升 15.23 个百分点、8.54 个百分点、5.23 个百分点。

基于 L-CVT、AlexNet、VGG16、ResNet18 的锑浮选工况识别的准确率对比曲线如图 12 所示。从图 12 可以看到，L-CVT 网络在验证过程中，识别准确率迅速提升，在 20 个验证轮数后基本稳定收敛，相较其他方法而言，L-CVT 网络的识别准确率曲线相对平滑稳定，且没有出现过拟合的现象，说明 L-CVT 网络具有良好的泛化能力和稳定的识别能力。

除此之外，表 6 显示了不同网络的计算复杂度。当输入浮选图像大小为 256×256、批量大小为 64 时，L-CVT 模型参数量为 2.38 MB，FLOPs 为 6.01×10^{10}。在相同实验环境及配置的情况下，所提模型相较于 AlexNet、VGG16、ResNet18 这 3 种模型而言：其参数量分别降低约 31.28 MB、43.91 MB、8.8 MB；计算成本降低约 3.32×10^{10}、5.08598×10^{12}、9.192×10^{10}。综上所述，L-CVT 网络具备更好的性能，能够更好地实现锑浮选工况的识别。

表 4　AlexNet、VGG16、ResNet18 的主要参数

Table 4　Main parameters of AlexNet, VGG16, ResNet18

AlexNet	VGG16	ResNet18
Layer-1: 11×11, 96; maxpool-3×3	Layer-1: $\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}$; maxpool-2×2	Layer-1: 7×7, 64 ; maxpool-3×3
Layer-2: 5×5, 96; maxpool-3×3	Layer-2: $\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}$; maxpool-2×2	Layer-2: $\begin{bmatrix}3\times3, 64\\3\times3, 64\end{bmatrix}\times3$
Layer-3: 3×3, 384	Layer-3: $\begin{bmatrix}3\times3, 256\\3\times3, 256\\3\times3, 256\end{bmatrix}$; maxpool-2×2	Layer-3: $\begin{bmatrix}3\times3, 128\\3\times3, 128\end{bmatrix}\times4$
Layer-4: 3×3, 384	Layer-4: $\begin{bmatrix}3\times3, 512\\3\times3, 512\\3\times3, 512\end{bmatrix}$; maxpool-2×2	Layer-4: $\begin{bmatrix}3\times3, 256\\3\times3, 256\end{bmatrix}\times6$
Layer-5: 3×3, 256; maxpool-3×3	Layer-5: $\begin{bmatrix}3\times3, 512\\3\times3, 512\\3\times3, 512\end{bmatrix}$; maxpool-2×2	Layer-5: $\begin{bmatrix}3\times3, 512\\3\times3, 512\end{bmatrix}\times3$
FC-1: 2048	FC-1: 4096	Pooling layer: average pool
FC-2: 2048	FC-2: 4096	FC-1: 6
FC-3: 6	FC-3: 6	Classifier: Softmax
Classifier: Softmax	Classifier: Softmax	

表 5　基于不同网络的锑浮选工况识别准确率

Table 5　Identificatiton accuracy of antimony flotation condition based on different networks

Network	L-CVT	AlexNet	VGG16	ResNet
Top-1 accuracy /%	93.56	78.33	85.11	88.33

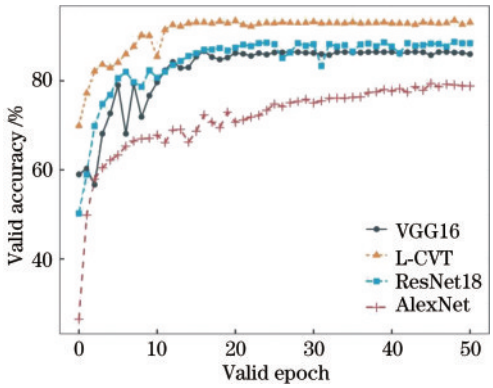

图 12　基于不同网络的锑浮选工况识别准确率对比曲线

Fig. 12 Comparison curves of identification accuracy of antimony flotation condition based on different networks

表 6　不同网络的计算复杂度

Table 6　Computational complexity of different networks

Network	Params /MB	FLOPs /10^9
L-CVT	2.38	60.10
AlexNet	33.66	93.31
VGG16	46.29	5146.08
ResNet18	11.18	152.02

为更全面地评估网络性能，采用混淆矩阵来反映 4 种网络对各个工况类别的实际识别情况，结果如图 13 所示。图 13 中：L-CVT 网络有 58 张浮选图片被错误识别，其余 842 张均正确识别；AlexNet 有 195 张浮选图片被错误识别，其余 705 张均正确识别；VGG16 有 134 张浮选图片被错误识别，其余 766 张均正确识别；ResNet18 有 105 张浮选图片被错误识别，其余 795 张浮选图片均正确识别。但所有网络均在第Ⅲ类和第Ⅳ类工况识别上出现较多误判，这是由于这 2 种工况类别在气泡形态上较为相似，再加上浮选工厂恶劣环境的影响，区分难度较大。但综合而言，所提模型对 6 种工况类别均有良好的识别能力，大多工况类别均能被正确识别。

通过混淆矩阵计算出每个网络的召回率、精确率、F1-Score 来作为评判模型识别效果的指标。为方便计算：定义预测结果为正、实际结果为正，则用 TP 表示；预测结果为正、实际结果为负，则用 FP 表示；预测结果为负、实际结果为正，则用 FN 表示；预测结果为负、实际结果为负，则用 TN 表示。在多类识别任务中，每个类别单独视为正，其余所有类别均视为负。计算出每个类别的精确率、召回率。为衡量整个网络的性能，再计算出 6 种工况类别的平均准确率和平均召回率来表示网络的准确率和召回率：

$$\begin{cases} P_i = \dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FP}}} \\ P = \mathrm{avg}\left(\sum_{i=1}^{6} P_i\right) \end{cases}, \tag{18}$$

$$\begin{cases} R_i = \dfrac{N_{\mathrm{TP}}}{N_{\mathrm{TP}} + N_{\mathrm{FN}}} \\ R = \mathrm{avg}\left(\sum_{i=1}^{6} R_i\right) \end{cases}。 \tag{19}$$

对模型而言，准确率和召回率是一对矛盾的度量。

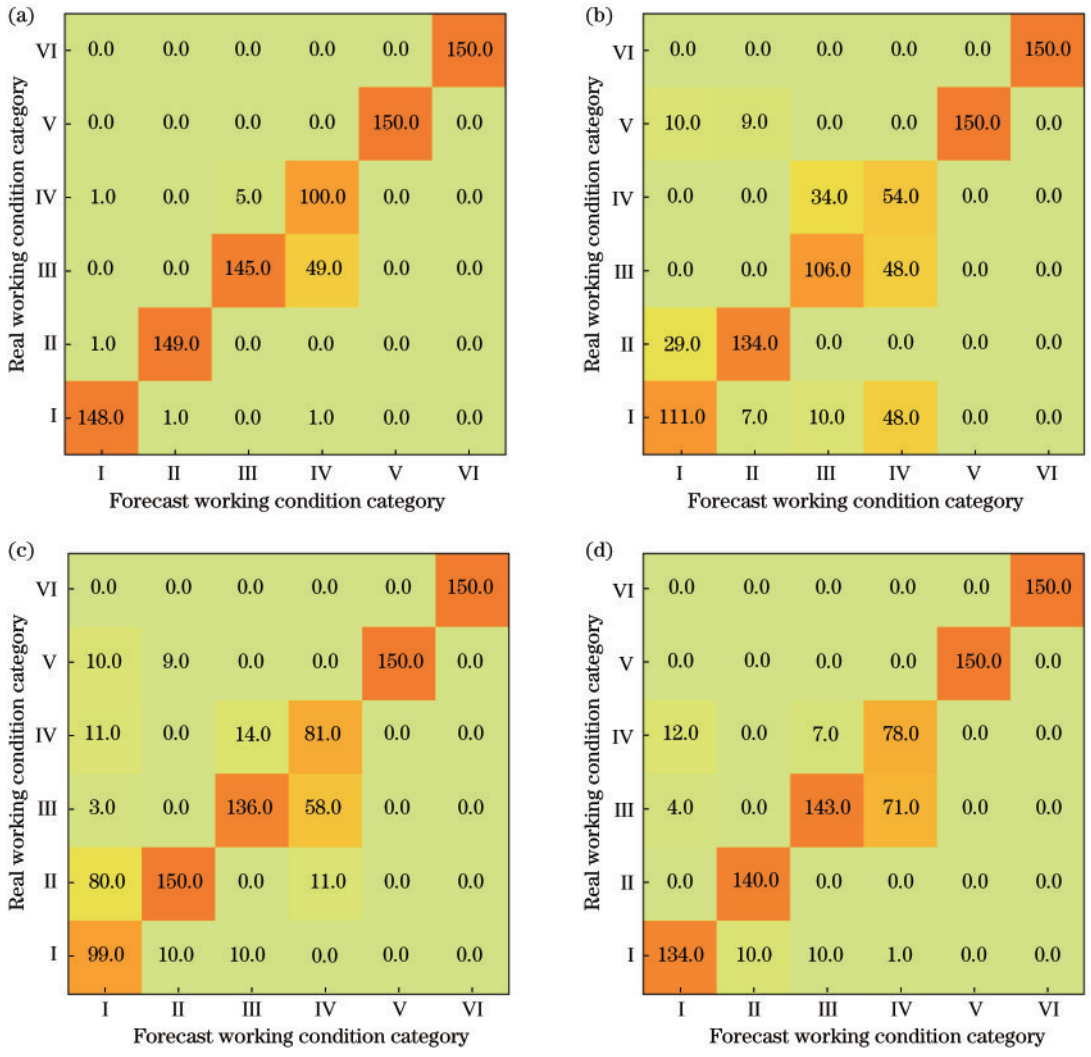

图 13　不同模型的混淆矩阵结果图。(a) L-CVT;(b) AlexNet;(c) VGG16;(d) ResNet18

Fig. 13　Confusion matrix results of different models. (a) L-CVT; (b) AlexNet; (c) VGG16; (d) ResNet18

通常情况下,召回率偏高时,精准率反而相对偏低;召回率偏低时,精确率反而相对偏高。为综合考量这两个指标,采用F1-Score来计算召回率和准确率的加权调和平均。F1-Score值越高,则表明该网络越稳定,性能更好,F1-Score的计算公式为

$$S_{F1}=\frac{2\times P\times R}{P+R}。\tag{20}$$

表7为不同网络的评估结果。从表7可以看到,基于L-CVT模型的锑浮选工况识别,相较于AlexNet、VGG16、ResNet18这3种工况识别而言:其精确率分别提升15.23个百分点、8.45个百分点、5.23个百分点;召回率提升17.14个百分点、9.05个百分点、4.77个百分点;其F1-Score分别提高16.18个百分点、8.75个百分点、5个百分点。即L-CVT网络进一步提高了锑浮选工况识别的准确性。

表7　不同网络的评估结果

Table 7　Evaluation results of different networks

Network	Precision /%	Recall /%	F1-Score /%
L-CVT	93.56	94.51	94.03
AlexNet	78.33	77.37	77.85
VGG16	85.11	85.46	85.28
ResNet18	88.33	89.74	89.03

为深入了解所提网络的性能相较于其他识别网络的优势,采用受试者工作特征(ROC)曲线可视化网络的性能,如图14所示,分别表示每个网络及其各个工况类别的ROC曲线和曲线下的面积(AUC)值。ROC曲线以false positive rate为横坐标,true positive rate为纵坐标,能较快地查找一个分类器在某个阈值对浮选图像的工况识别能力,曲线越靠近左上角,则误判率越低、灵敏度越高。AUC表示ROC曲线下的面积,用于衡量模型的泛化能力,反映了ROC曲线表达的分类能力。AUC越高,则模型的工况识别效果就越好。结合ROC曲线和AUC可以看出,L-CVT网络相较于AlexNet、VGG16、ResNet18等网络对锑浮选工况的识别效果更好、抗干扰能力更强,具有更好的鲁棒性和泛化性。

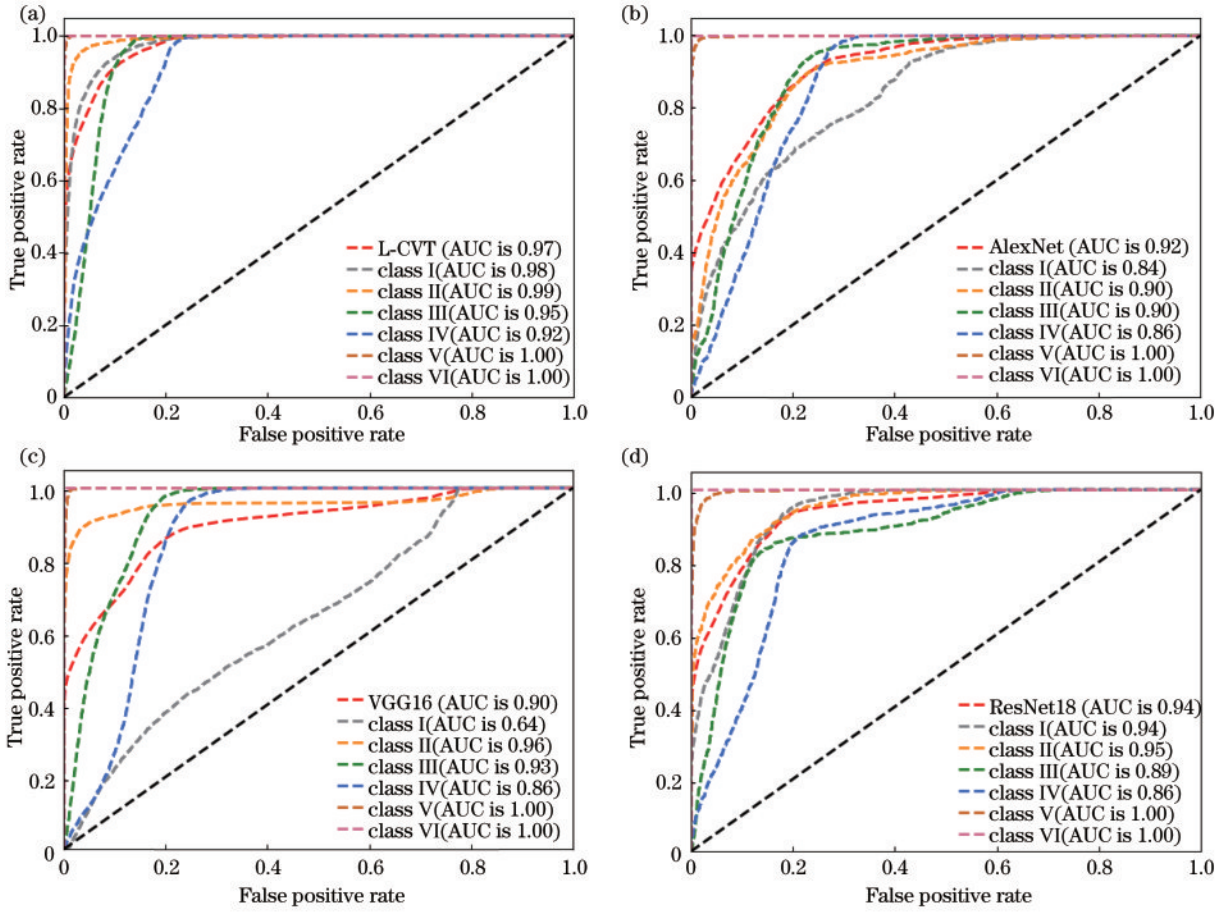

图 14　不同网络的 ROC 曲线和 AUC 值。(a)L-CVT;(b)AlexNet;(c)VGG16;(d)ResNet18

Fig. 14　ROC curves and AUC values of different networks. (a) L-CVT; (b) AlexNet; (c) VGG16; (d) ResNet18

4.5　消融实验

为进一步评估 L-CVT 网络中各模块性能，本小节实验通过删除和替换各重要组成模块，以确定各模块对网络性能的影响。在所提方法上，删除 Conv-VIT 模块，并使用 3×3 的标准卷积核替代 L-Conv 模块来执行卷积，将此作为基础方法。然后比较：1)基础方法的识别准确率；2)在基础方法上，加入 Conv-VIT 模块后的识别准确率；3)将基础方法中标准卷积核替换为 L-Conv 模块后的识别准确率；4)在基础方法上，同时加入 2)、3)操作后的识别准确率。实验结果如表 8 所示。

表 8　消融实验

Table 8　Ablation experiment

Method	Accuracy /%	F1-score /%
Base	75.22	74.60
Base + A	88.78	88.70
Base + B	81.44	81.70
Base + A + B (proposed network)	93.56	94.03

表 8 中：Base 表示基础方法；A 表示在基础方法上加入 Conv-VIT 模块；B 表示将基础方法中标准卷积核替换为 L-Conv 模块。从表 8 可以看到，在基础方法中加入 A 方法比加入 B 方法的识别准确率更高、模型性能更好，说明 Conv-VIT 模块对网络性能影响较大，产生较好的工况识别效果。此外，通过融合 L-Conv 模块和 Conv-VIT 模块发现，所提方法具有较好的工况识别准确率。

4.6　网络的特征图可视化

为进一步了解所提方法如何提取浮选图像特征，采用 Grad-CAM 进行特征可视化。该算法通过平均梯度加权 2D 激活的方式来获取分类权重。最后，将导向反向传播和热力图逐点相乘的方式来获得高分辨率的 Grad-CAM 可视化图。Grad-CAM 可描述为

$$\begin{cases} \partial_k^c = \dfrac{1}{N}\sum_i \sum_j \dfrac{\partial y^c}{\partial A_{ij}^k} \\ M_{\mathrm{Grad-CAM}}^c = \mathrm{ReLU}\left(\sum_k \partial_k^c A^k\right) \end{cases}, \tag{21}$$

式中：N 是浮选泡图中的像素数目；y^c 是对应类 c 的分数；∂A_{ij}^k 表示浮选图 A^k 在位置 (i,j) 处的激活。最后加入 ReLU，使得只关注对类别 c 有正向影响的像素点，同时也避免代入其他类别的像素而影响最终解释效果的风险。

从浮选数据集中随机选择几张锑浮选图片，使用 Grad-CAM 算法可视化 L-CVT、AlexNet、ResNet18、

VGG16这4种不同的模型。特征表示为最后一层分类层的输出，如图15所示。图15中底下的JET颜色条显示了分级权重，不同的颜色代表着不同的权重值，深蓝色表示低分类权重，亮红色表示高分类权重，颜色变化由深蓝色渐变到亮红色，所对应的权重值由小到大，权重值的高低代表模型学习特征能力的强弱。通过对比这4种网络的热力图发现，它们的特征提取方式截然不同，L-CVT相较于其余3种网络而言，其权重值的分布更广泛，即关注的区域信息更全面，说明L-CVT网络具备显著的全局视野，能够远距离捕获浮选泡沫的特征。而其余网络的权重值分布大多集中于部分区域，会忽视浮选泡沫图像内部的长距离依赖关系。本小节实验很好地解释了L-CVT网络具备全局建模能力的原因。

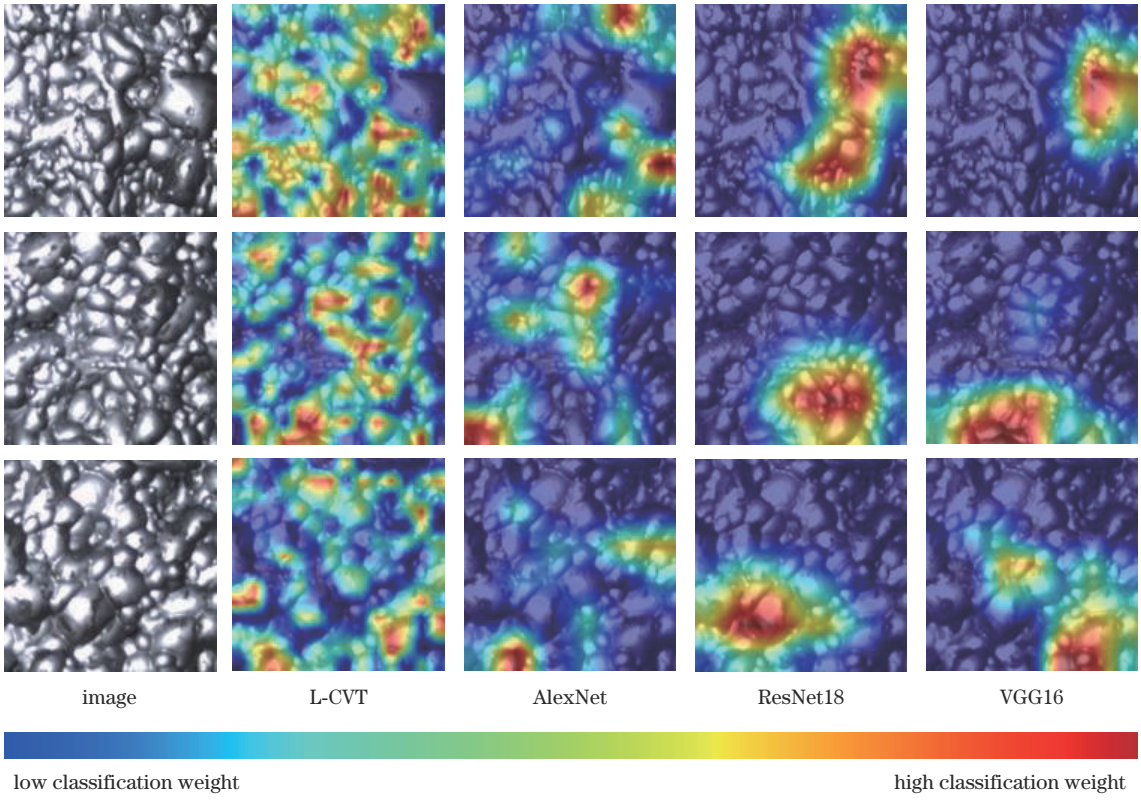

图15　4种网络的特征图可视化结果

Fig. 15　Visualization results of feature maps of four kinds of networks

5　结　　论

针对锑浮选工况识别误差大、效率低等问题，提出的基于L-CVT网络的锑浮选工况识别技术，解决传统CNN会忽略浮选泡沫图像内部长距离依赖关系、无法准确描述浮选泡沫表面的复杂特征信息等问题，从而提高浮选工况识别的准确率。同时引入轻量型CNN模块，减少因Transformer层执行注意力操作带来的计算成本，在实际工业中节省大量的计算资源。经实验比对，所提方法能在恶劣的浮选环境下高效地识别锑浮选的工况类别，也为实现浮选工况自动化生产奠定基础。

将深度学习方法运用到工业锑浮选工况的识别中，取得较好的识别效果，有效地解决人工观察浮选泡沫无法客观评价浮选工况状态的问题，为锑浮选工况的识别提供很好的理论支持，但仍存在一定的局限性。在实际浮选现场中客观存在环境较为恶劣、光照不均等问题，导致所提方法对于部分噪声过多、细节信息丢失严重的浮选图像的识别效果并不理想。下一阶段将重点针对这一问题进行深入研究，提升浮选图像预处理方法效果，减少恶劣环境带来的图像噪声，提高锑浮选工况识别的准确率，使所提方法能更好地运用于实际工业环境中。

参　考　文　献

[1] Quintanilla P, Neethling S J, Brito-Parada P R. Modelling for froth flotation control: a review[J]. Minerals Engineering, 2021, 162: 106718.

[2] 桂卫华, 阳春华, 徐德刚, 等. 基于机器视觉的矿物浮选过程监控技术研究进展[J]. 自动化学报, 2013, 39(11): 1879-1888.
Gui W H, Yang C H, Xu D G, et al. Machine-vision-based online measuring and controlling technologies for mineral flotation: a review[J]. Acta Automatica Sinica, 2013, 39(11): 1879-1888.

[3] Bartolacci G, Pelletier P, Jr, Tessier J, Jr, et al. Application of numerical image analysis to process diagnosis and physical parameter measurement in mineral

processes: part I: flotation control based on froth textural characteristics[J]. Minerals Engineering, 2006, 19(6/7/8): 734-747.

[4] Ai M X, Xie Y F, Xu D G, et al. Data-driven flotation reagent changing evaluation via union distribution analysis of bubble size and shape[J]. The Canadian Journal of Chemical Engineering, 2018, 96(12): 2616-2626.

[5] 阳春华, 任会峰, 桂卫华, 等. 基于泡沫纹理信度分配SVM的矿物浮选工况识别[J]. 仪器仪表学报, 2011, 32(10): 2205-2209.
Yang C H, Ren H F, Gui W H, et al. Performance recognition using texture credit distributed SVM for froth flotation process[J]. Chinese Journal of Scientific Instrument, 2011, 32(10): 2205-2209.

[6] 梁秀满, 田童, 刘文涛, 等. 基于泡沫图像特征融合的煤泥浮选工况识别[J]. 计算机仿真, 2021, 38(4): 385-389.
Liang X M, Tian T, Liu W T, et al. Coal slime flotation condition identification based on fusion of froth image features[J]. Computer Simulation, 2021, 38(4): 385-389.

[7] Ai M X, Xie Y F, Tang Z H, et al. Deep learning feature-based setpoint generation and optimal control for flotation processes[J]. Information Sciences, 2021, 578: 644-658.

[8] Fu Y H, Aldrich C. Froth image analysis by use of transfer learning and convolutional neural networks[J]. Minerals Engineering, 2018, 115: 68-78.

[9] 张进, 廖一鹏, 陈诗媛, 等. 基于多尺度CNN特征及RAE-KELM的浮选加药状态识别[J]. 激光与光电子学进展, 2021, 58(12): 1215002.
Zhang J, Liao Y P, Chen S Y, et al. Flotation dosing state recognition based on multiscale CNN features and RAE-KELM[J]. Laser & Optoelectronics Progress, 2021, 58(12): 1215002.

[10] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale[EB/OL]. (2020-10-22) [2021-02-05]. https://arxiv.org/abs/2010.11929.

[11] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, December 4-9, 2017, Long Beach, CA, USA. New York: ACM Press, 2017: 6000-6010.

[12] Sandler M, Howard A, Zhu M L, et al. MobileNetV2: inverted residuals and linear bottlenecks[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4510-4520.

[13] Xu P C, Zhao J X, Zhang J. Identification of intrinsically disordered protein regions based on deep neural network-VGG16[J]. Algorithms, 2021, 14(4): 107-113.

[14] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[15] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[16] Selvaraju R R, Cogswell M, Das A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 618-626.

[17] 曹斌芳, 谢永芳, 桂卫华, 等. 锑浮选过程粗选扫选工序的药剂量协调优化方法[J]. 中南大学学报(英文版), 2018, 25(1): 95-106.
Cao B F, Xie Y F, Gui W H, et al. Coordinated optimization setting of reagent dosages in roughing-scavenging process of antimony flotation[J]. Journal of Central South University, 2018, 25(1): 95-106.

[18] Ai M X, Xie Y F, Xie S W, et al. Fuzzy association rule-based set-point adaptive optimization and control for the flotation process[J]. Neural Computing and Applications, 2020, 32(17): 14019-14029.

[19] 申毫, 孟庆浩, 刘胤伯. 基于轻量卷积网络多层特征融合的人脸表情识别[J]. 激光与光电子学进展, 2021, 58(6): 0610005.
Shen H, Meng Q H, Liu Y B. Facial expression recognition by merging multilayer features of lightweight convolutional networks[J]. Laser & Optoelectronics Progress, 2021, 58(6): 0610005.

[20] Ioffe S, Szegedy C. Batch normalization: accelerating deep network training by reducing internal covariate shift[C]//32nd International Conference on Machine Learning, ICML 2015, July 6-11, 2015, Lille, France. Cambridge: JMLR, 2015: 448-456.

[21] 张家钧, 唐云祁, 杨智雄, 等. 基于注意力机制的鞋型识别算法[J]. 激光与光电子学进展, 2022, 59(2): 0215004.
Zhang J J, Tang Y Q, Yang Z X, et al. Shoe type recognition algorithm based on attention mechanism[J]. Laser & Optoelectronics Progress, 2022, 59(2): 0215004.

[22] Srivastava N, Hinton G E, Krizhevsky A, et al. Dropout: a simple way to prevent neural networks from overfitting[J]. Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[23] Garg B, Sharma G K. A quality-aware energy-scalable gaussian smoothing filter for image processing applications[J]. Microprocessors and Microsystems, 2016, 45: 1-9.

[24] Jindal A, Gnaneshwar D, Sawhney R, et al. Leveraging BERT with mixup for sentence classification (student abstract)[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(10): 13829-13830.

[25] Yun S, Han D, Chun S, et al. CutMix: regularization strategy to train strong classifiers with localizable features[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 6022-6031.

[26] 陆雅诺, 陈炳才, 陈德刚, 等. 一种基于注意力模型的带钢表面缺陷识别算法[J]. 激光与光电子学进展, 2021, 58(14): 1410014.
Lu Y N, Chen B C, Chen D G, et al. Recognition algorithm of strip steel surface defects based on attention model[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410014.