

细粒度鱼类疫病图像识别算法模型

魏立明^{1*}, 赵奎¹, 王宁¹, 张忠岩², 崔海朋²

¹中国海洋大学信息科学与工程学部, 山东 青岛 266100;

²青岛杰瑞工控技术有限公司, 山东 青岛 266071

摘要 通过肉眼识别鱼类疫病依赖于诊断人员的经验,疫病数据存在类间差距较小与识别效率低等细粒度问题。由于 Transformer 缺乏卷积神经网络(CNN)的归纳偏差,需要大量的数据进行训练;CNN 对全局特征提取不足,泛化性能较差等问题限制模型的分类精度。基于特征图对所有像素的全局交互建立算法模型,提出一种基于 CNN 与 Vision Transformer 相结合的鱼类疫病识别模型(CViT-FDRM)。首先,搭建鱼类疫病的数据库 FishData01;其次,利用 CNN 提取鱼类图像细粒度特征,采用 Transformer 模型自注意力机制获取图像全局信息进行并行训练;然后,采用组归一化层将样本通道分组求均值与标准差;最后,采用 404 张鱼类疫病图像进行测试,CViT-FDRM 达到 97.02% 的识别准确率。在细粒度图像开源数据库 Oxford Flowers 上的实验结果表明,CViT-FDRM 的分类精度优于主流的细粒度图像分类算法,可达 95.42%,提高 4.84 个百分点。CViT-FDRM 在细粒度图像识别方面可达到较好的效果。

关键词 图像处理; 卷积神经网络; Vision Transformer; 细粒度; 鱼类疫病识别; CViT-FDRM

中图分类号 TP399

文献标志码 A

DOI: 10.3788/LOP222630

Fine-Grained Fish Disease Image Recognition Algorithm Model

Wei Liming^{1*}, Zhao Kui¹, Wang Ning¹, Zhang Zhongyan², Cui Haipeng²

¹Department of Information Science and Engineering, Ocean University of China, Qingdao 266100, Shandong, China;

²Qingdao JARI Industrial Control Technology Co., Ltd., Qingdao 266071, Shandong, China

Abstract Identification of fish epidemics by the naked eye depends on the experience of diagnostic personnel. Moreover, the epidemic data has such fine granularity problems as small gaps between categories and low recognition efficiency. Usually, the Transformer requires a large amount of data for training due to the lack of inductive bias in convolutional neural networks (CNN). In addition, the model's classification accuracy is restricted by insufficient global feature extraction and the weak generalization performance of CNN. In this study, based on the global interaction of all pixels in the feature map, an algorithm model is developed, and a fish epidemic recognition model (CViT-FDRM) using the combination of CNN and a Vision Transformer is suggested. First, FishData01, a database of fish epidemics, is set up. Second, CNN is used to extract the fine-grain features of fish images, and the Transformer model self-attention mechanism is used to acquire the global information of images for parallel training. Then, the group normalization layer is utilized to group the sample channels to compute the mean and standard deviation. Finally, 404 fish epidemic images were used for testing, and CViT-FDRM obtained 97.02% recognition accuracy. The experimental results on Oxford Flowers, an open-source database of fine-grained images, reveal that CViT-FDRM has greater classification accuracy than that of the standard fine-grained image classification algorithm, reaching 95.42%, which is 4.84 percentage points higher. Therefore, CViT-FDRM can perform well in fine-grain image recognition.

Key words image processing; convolutional neural network; Vision Transformer; fine grit; fish disease recognition; CViT-FDRM

收稿日期: 2022-09-26; 修回日期: 2022-10-25; 录用日期: 2022-11-23; 网络首发日期: 2023-01-04

基金项目: 山东省重点研发计划(科技示范工程)(2021SFGC0701)

通信作者: *15865569879@163.com

1 引言

推进现代化渔业发展,实现海洋强国战略是我国重要发展方针。渔业活动为人类提供丰富的生活必需品的同时面临着各种疫病威胁^[1]。目前,鱼类健康监测仍采用传统的方法,但人类可视化识别方式准确性低、耗时长、容易产生各种各样误报^[2-3]。因此,在养殖经验和鱼类健康专业知识外,需要快速可靠的鱼类疫病识别工具^[4-5]。计算机视觉方法在不影响输出精度的情况下自动提取特征,在鱼类疫病识别中发挥重要作用^[6]。

在细粒度图像分析(FGIA)中,细粒度特性导致类间差距较小和类内差距较大,使其成为一个具有挑战性的问题。FGIA 的任务是从下属类别中准确识别出物体的种类,FGIA 与通用图像识别与分析的不同之处在于:前者对象是来自元类别的子类别,后者的目标是粗粒度的元类别,因此前者对细粒度的相似特征进行识别较为困难。

Waleed 等^[7]提出自动检测养殖场流行性溃疡综合征、鱼鳞病和柱状线虫病的方法,在不同颜色空间中使用不同卷积神经网络(CNN)架构进行细粒度识别。双线性 CNN 方法将提取的特征值相乘增强识别能力从而实现鱼类疫病细粒度识别,但难以大规模应用^[8]。针对这一问题,Gao 等^[9-11]通过应用张量草图进行低维嵌入,保持更高的识别精度,在前人研究的基础上通过设计专属的细粒度图像损失函数实现细粒度的分类,提高模型的分类精度。

Transformer 和自注意力机制模型极大地促进了机器学习发展^[12]。最初 Transformer 用于处理由 CNN 骨干提取的图像序列特征^[13]。随着不断的深入研究和不断发展,Transformer 模型进一步扩展到其他主流的计算机视觉任务中^[14-20]。Vision Transformer(ViT)^[21]是视觉 Transformer 中备受关注之作,将 Transformer 直接应用于图像序列中可以达到很好的图像分类效果。He 等^[22-23]提出的 TransReID 在 Transformer 中嵌入信息,提高目标再识别性能。

综上所述,图像识别技术可以用最少的时间在不影响输出精度的情况下自动提取相关特征,分析并完成细粒度鱼类疫病识别。本研究的主要贡献如下:1) 在不同水域、不同季节、不同天气下与养殖专家开展数据采集工作,并形成细粒度鱼类疫病识别数据集 FishData01。通过图像处理技术提高模型在复杂多变环境下的抗干扰能力,提高模型的泛化能力。2) 基于图像识别的鱼类疫病检测现状,利用 CNN 处理图像的优势提取图像特征,利用 Transformer 模型自注意力机制,对图像的不同位置提供不同的注意捕捉网络每个阶段的局部和全局空间相互作用。充分结合 CNN 和 Transformer 的优势提出细粒度鱼类疫病识别模型(CViT-FDRM)。3) 采用组归一化(GN)层替代原有

层归一化(LN)层,将样本通道分组求均值与标准差,提高鱼类疫病识别准确率。

2 CNN-Vision Transformer 融合网络

2.1 CViT-FDRM 网络的整体结构

所提鱼类疫病识别模型 CViT-FDRM 如图 1 所示,该架构主要包含卷积层、embedding 层、Transformer encoder 层与 multi-layer perceptron(MLP) head 层。

CViT-FDRM 采用混合的 CNN-Transformer 架构,包括卷积结构和 Transformer 结构,利用来自 CNN 的特征,保留更多的图像边缘信息。卷积运算的局部特性使得距离较远的特征之间难以沟通,导致提取的特征中含有背景复杂的部分,影响识别结果。通过在卷积结构之后连接 Transformer 结构来解决这个问题。与以往基于 CNN 的自我注意方法相比,基于 Transformer 的模型架构提供更好的自我注意方法——Transformer 全局注意力机制,Transformer 中的多头注意机制可以建立对输入图像的长距离依赖,对图像的不同位置提供不同的注意,从而实现精确特征提取与鱼类疫病识别。

2.2 卷积神经网络特征提取

2.2.1 卷积层

CNN 是由执行不同任务的独立块连接起来组成的。在 CNN 中,这些构建模块被称为层。利用 CNN 对鱼类图像进行特征提取时,图像通过输入层输入网络中,然后对其进行归一化处理。图像高度、宽度和通道大小分别设置为 224、224、3(红、绿、蓝)。卷积层以空间感知的方式将各种低层特征编码为更具有辨析性的特征,卷积层可以被认为是一系列过滤器,将输入图像转换为其他特征模式。使用卷积层从输入鱼类图像中提取特征。然后将滤波后的原始像素值图像送入 RGB 颜色通道,测量滤波像素与输入像素之间的点积。将提取图像某些特征的权值矩阵设为 3×3 矩阵,CNN 在整个训练阶段学习过滤器的值。卷积层之后是非线性层,可以用来修改或禁用输出,这一层用于饱和或限制产生输出。采用修正线性单元(ReLU),提供更简单的函数和梯度定义。输入的元素都要接受一个阈值操作,其中任何小于零的值被设为零。

最后通过自注意力机制模块集成相关特征进行图像分类,通过融合全局信息达到准确线性分类的目的。模型中的各个节点与上一层和下一层中的各个节点都有唯一的关系。由于这个过程涉及空间信息的丢失,并且是不可逆的,输出由加起来等于 1 的正值组成,这一层将每个输入分配给一个互斥类,并利用 Softmax 激活函数提供的概率来计算损失,分类层可以将其用作分类概率。分类是将一幅图像按其所属的鱼类疾病类型进行分类的过程。

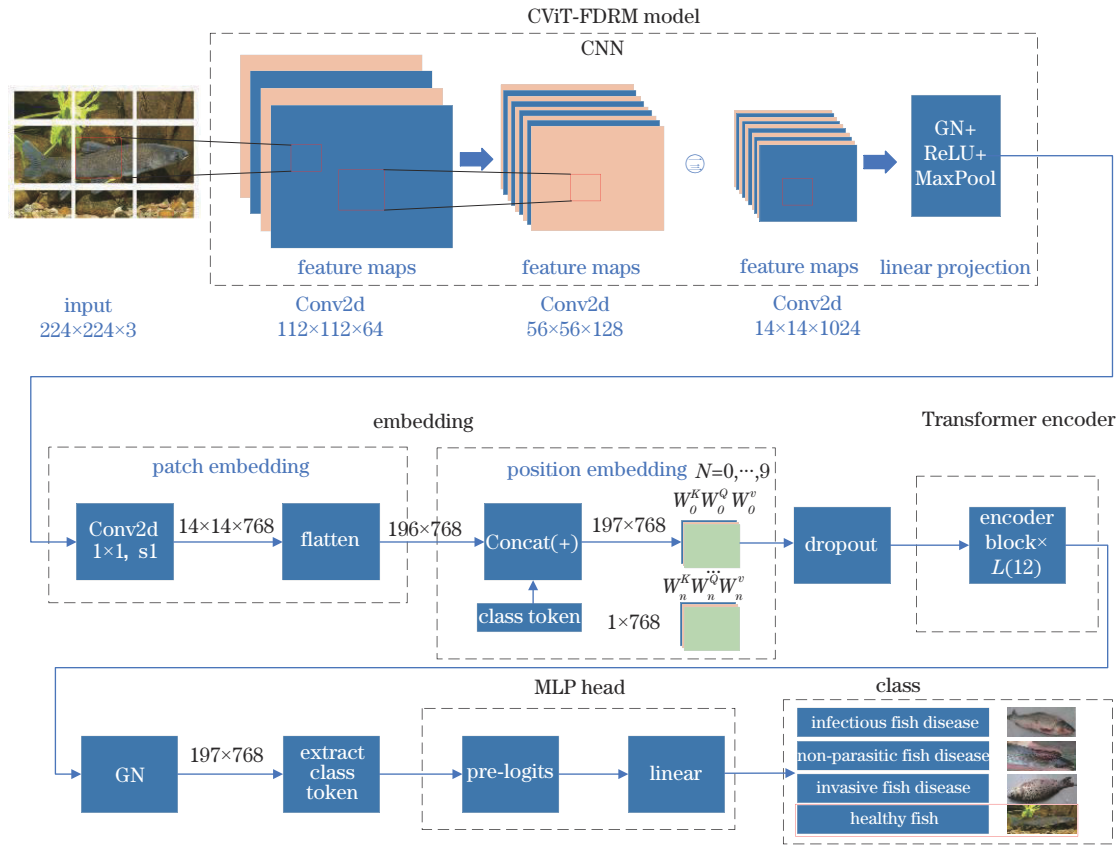


图 1 细粒度鱼类疫病识别模型 CViT-FDRM

Fig. 1 Fine grained fish disease identification model CViT-FDRM

2.3 Vision Transformer 网络

2.3.1 linear projection of flattened patches (embedding)

标准的 Transformer 模块如图 2 所示。输入 Transformer encoder 中要求的是二维矩阵向量,本研究

中每个向量的长度为 768。鱼类疫病的数据集为 $[H, W, C]$ 的图像数据,这种类型数据格式不是 Transformer 所要求的二维数据,因此需要增加 linear projection of flattened patches,简称 embedding 层。图 2

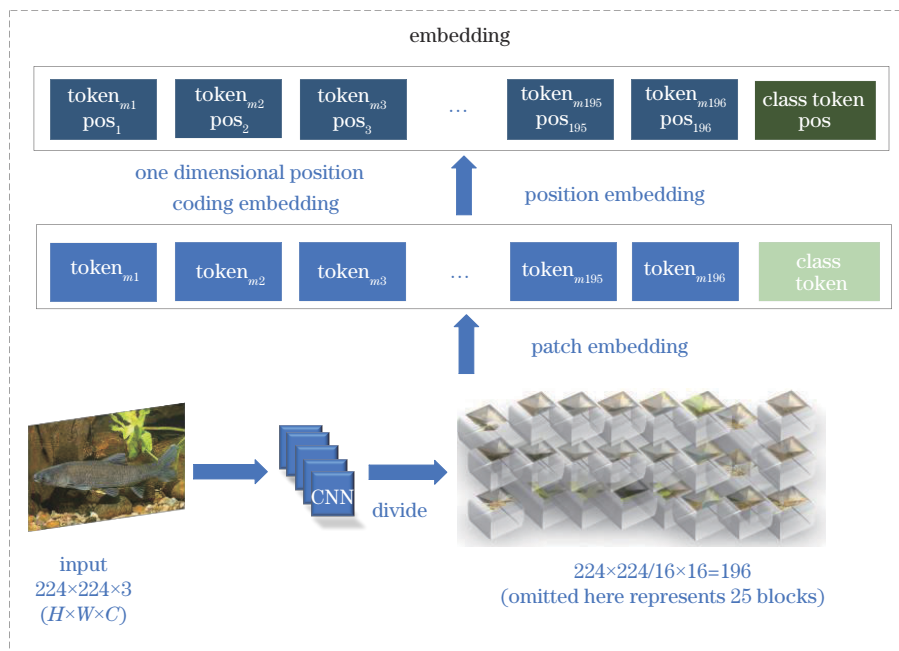


图 2 CViT-FDRM 模型的 embedding 层

Fig. 2 embedding layer of CViT FDRM model

中,将鱼类图像数据分为若干个相同大小的块,224×224 的图片按照 16×16 大小进行划分得到 196[(224×224)/(16×16)]个块。接着将每个[16, 16, 3]的块通过线性映射生成长度为 768 的一维向量。

鱼类疫病图像识别中使用卷积核大小为 16×16、卷积核个数为 768 的卷积层来实现。通过卷积[224, 224, 3]生成[14, 14, 768],将宽、高两个维度展平成[196, 768]的二维矩阵并输入 Transformer 模型中。同时添加类别向量与 position embedding。与之前生成的二维矩阵叠加,196 个块中需添加作为分类的额外块,因此分块的总数量为 197,由([1, 768], [196, 768])生成 [197, 768]。position embedding 即 positional encoding,确定不同块的位置信息,提高时序序列捕获能力,提高识别准确率。

$$P_{p,2i} = \sin(p/10000^{2i, d_{\text{model}}}), \quad (1)$$

$$P_{p,2i+1} = \cos(p/10000^{2i, d_{\text{model}}}), \quad (2)$$

式中: $P_{p,2i}$ 、 $P_{p,2i+1}$ 是位置矩阵中的元素; p 表示向量的某一具体位置, $p \in [0, \max(\text{length})]$; d_{model} 表示维度,具体某一个维度 $i \in [0, d_{\text{model}}/2]$ 。

2.3.2 Transformer encoder 层

鱼类疫病图像数据通过 embedding 层后输入 Transformer encoder 层中,该层重复多次实现 encoder block 操作。

由于学习不同类别数据集的分布时,没有进行归一化处理的数据分布不统一,会增大训练成本。LN 与实例归一化(IN)可理解为 GN 的一种特例,因此合理地选择组大小可以达到更好的分类效果,采用 GN 层对每个向量进行特征归一化处理,减少数据分布的差异。

对输入的特征信息进行组归一化,计算公式为

$$\mu = \frac{1}{B \times H \times W} = \sum_{h=1}^H \sum_{w=1}^W \sum_{k=(G-1) \times S+1}^{G \times S} a, \quad (3)$$

$$\sigma^2 = \frac{1}{B \times H \times W} \sum_{h=1}^H \sum_{w=1}^W \sum_{k=(G-1) \times S+1}^{G \times S} (a - \mu)^2, \quad (4)$$

$$sm = \frac{a - \mu}{\sigma}, \quad (5)$$

$$n = \gamma m + \beta, \quad (6)$$

式中: a 为输入特征值; B 是输入的批处理量; W 与 H 为输入的特征的宽和高; μ 为输入特征的均值; σ 为输入特征的方差; C 为输入的通道数; G 为组归一化处理输入数据所分成的组数量;每组所分得的通道数量 $S=C/G$, S 的值小于等于通道数量。输入特征数据减去均值并除方差进行归一化,为了确保模型的学习能力,引入比例参数 γ 与位移参数 β ,由此可得出组归一化处理后的特征数据。

multi-head attention 基于特征图对所有像素的全局交互建立算法模型,对不同序列进行注意力运算。加强多个头部相互作用的学习,头部能够共享学习,将

归一化后的特征输入多头注意力机制中。

$$\begin{cases} \mathbf{q}^m = \mathbf{W}^q \mathbf{x}^m \\ \mathbf{k}^m = \mathbf{W}^k \mathbf{x}^m, \\ \mathbf{v}^m = \mathbf{W}^v \mathbf{x}^m \end{cases} \quad (7)$$

$$\begin{cases} \mathbf{q}^{m,1} = \mathbf{W}^{q,1} \mathbf{q}^m \\ \mathbf{k}^{m,1} = \mathbf{W}^{k,1} \mathbf{k}^m, \\ \mathbf{v}^{m,1} = \mathbf{W}^{v,1} \mathbf{v}^m \end{cases} \quad (8)$$

式中: \mathbf{q}^m 、 \mathbf{k}^m 、 \mathbf{v}^m 分别是输入序列 \mathbf{x} 与各自权重矩阵 \mathbf{W} 相乘所得; q 、 k 、 v 分别表示 query(匹配其他)、key(待匹配)、value(待提取信息)。权重矩阵再与变换矩阵相乘得 $\mathbf{q}^{m,1}$ 、 $\mathbf{k}^{m,1}$ 、 $\mathbf{v}^{m,1}$ 。 $\mathbf{q}^{m,1}$ 与 $\mathbf{k}^{m,1}$ 做一定的点积运算得到 $\mathbf{v}^{m,1}$ 的权重矩阵,然后与 $\mathbf{v}^{m,1}$ 相乘; $\mathbf{q}^{m,1}$ 与 $\mathbf{k}^{n,1}$ 做一定的点积运算得到 $\mathbf{v}^{n,1}$ 的权重矩阵,然后与 $\mathbf{v}^{n,1}$ 相乘。最终将所得的结果叠加得到 $\mathbf{b}^{m,1}$,同理可得 $\mathbf{b}^{m,2}$ 。 $\mathbf{b}^{m,1}$ 、 $\mathbf{b}^{m,2}$ 与变换矩阵相乘之后通过叠加操作得到多头注意力最终输出。

$$\text{Multihead} = \text{Contact}(\text{head}_0, \text{head}_1, \dots, \text{head}_n). \quad (9)$$

式中:head为注意力机制各个头所得的结果;Contact(\cdot)代表叠加运算;Multihead代表多头注意力机制,将多个头叠加得到最终输出。给输入图像的不同区域提供不同的注意力,通过分配不同的权重,最后对特征图进行叠加生成能分辨不同疫病类别的特征图,分类可视化示意图如图 3 所示。

MLP block 由 linear、GELU 激活函数与 dropout 组成,如图 4 所示。linear 将输入的节点数量由[197, 768]扩充到[197, 3072],相比较原节点数量扩充了 4 倍,下一个 linear 层将节点的个数由[197, 3072]还原到[197, 768],输入 dropout 层中。

实际训练过程中模型的训练样本可能不够,数据集会存在不平衡等问题,为了防止模型训练需要的参数过多出现训练集精度较高,而测试集精度较低的过拟合问题,在网络的最后接入 dropout 层。

2.3.3 MLP head 层

MLP head 层主要由 linear 与激活函数 GELU 组成,通过 Transformer encoder 后数据的输入 shape 和输出 shape 是保持不变的,例如输入[197, 768],则输出[197, 768]。提取所需的类别向量,就是从[197, 768]向量中抽取出类别对应的[1, 768]向量,之后通过 MLP head 层得到鱼类疫病的最终分类结果。

2.4 算法工作流程

1)将任意鱼类图像输入 CViT-FDRM 中进行细粒度鱼类疫病识别,然后通过图像处理技术的序列化对图像进行细化,将处理后的数据集通过 CNN 提取图像特征。

2)将处理后的数据输入 embedding 层进行 patch embedding 与 position embedding 处理,将输入的图像分割成小 patch,转化为一维并嵌入一维位置信息便于进行后续操作。

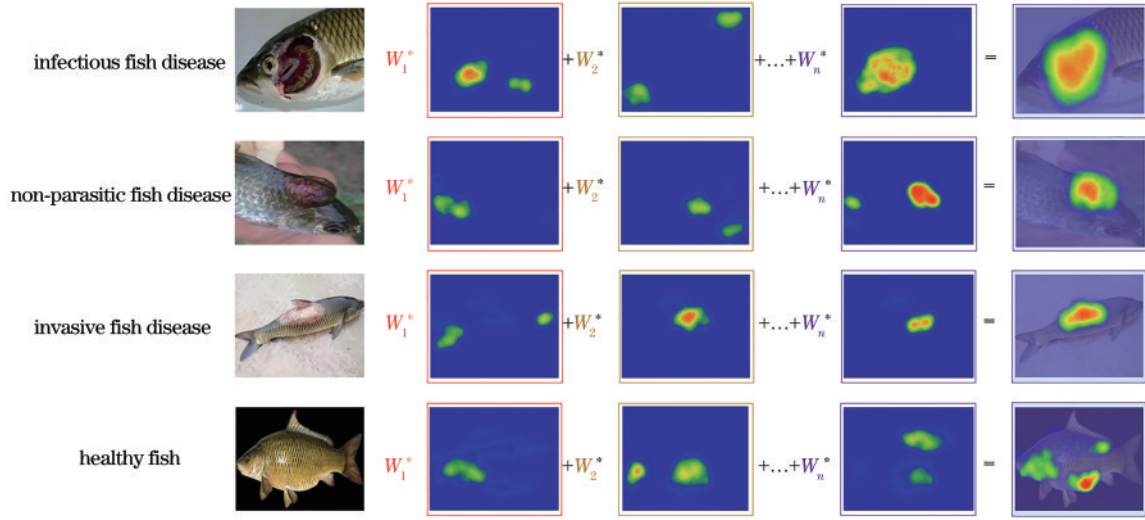


图 3 模型分类可视化

Fig. 3 Model classification visualization

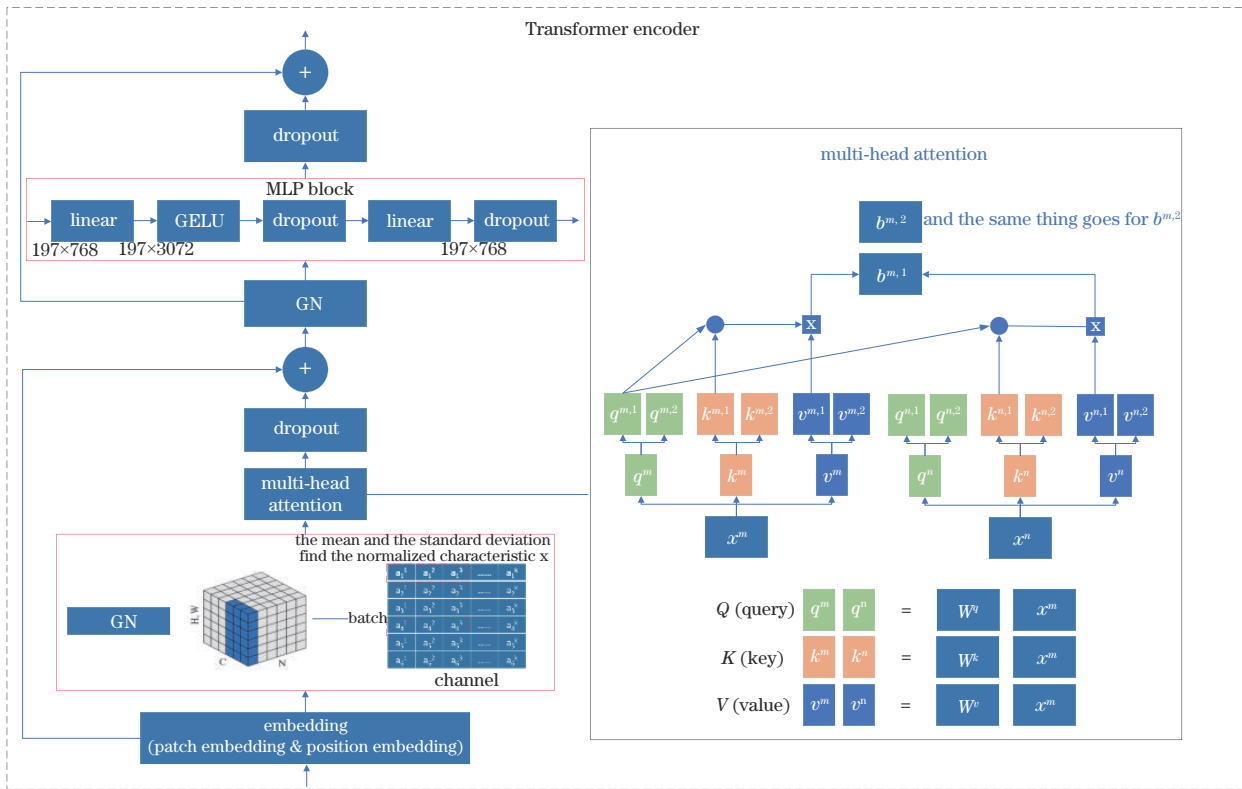


图 4 CViT-FDRM模型的Transformer encoder层

Fig. 4 Transformer encoder layer of CViT-FDRM model

3) 将 embedding 层处理后的数据通过 Transformer encoder 层, 采用组归一化操作, 对比选择最合适的组数量, 通过多次组归一化与多头注意力机制后进行线性连接。MLP head 输出的结果是输入图像的标签, 用于将特定类别分类为健康鱼或者感染的鱼。具体工作流程如图 5 所示。

3 实验结果分析

3.1 数据集

不同季节的水温和受污染的水都是鱼类疾病传播的重要因素, 有些疾病呈现复杂可见模式, 通过详细分析掌握鱼类疾病尤为重要。首先, 必须认识到鱼类疫

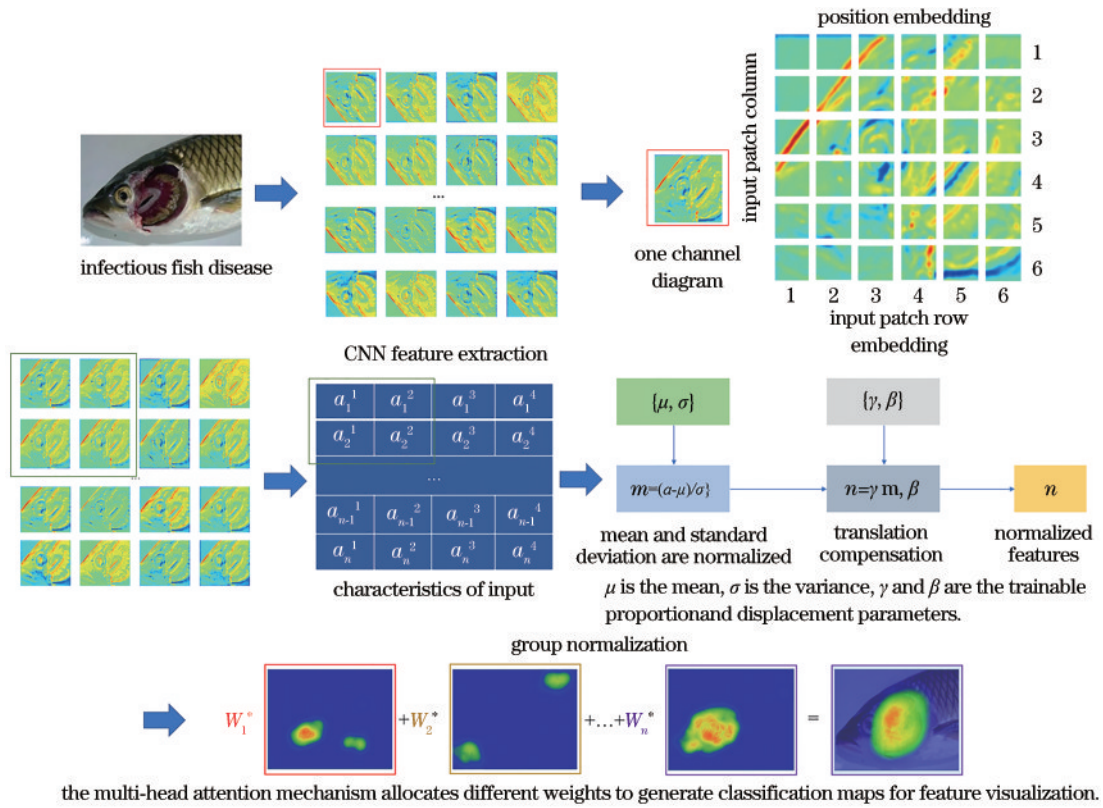


图 5 算法计算过程可视化

Fig. 5 Visualization of algorithms calculation process

病识别中的问题;然后,进一步研究基于计算机框架下处理问题的潜在策略。为排除鱼病,可对感染鱼的皮肤、鳃、眼睛或鳞片上的可见痕迹进行识别,包括病变、侵蚀、割伤、红肿与肿块^[24-25]。细粒度鱼类疫病图像如

图 6 所示,对多个相似度很高的病种进行识别,如传染性鱼病、非寄生引起的鱼病、侵袭性鱼病与健康鱼类等。为了实现准确识别,需要捕捉细微的差异(如鱼眼、鱼鳃与鱼尾等)来区分。



图 6 鱼疫病状态示意图

Fig. 6 Schematic diagrams of fish epidemic disease

为提高鱼类疫病数据集的多样性,分别在不同水域、不同季节、不同天气下与养殖专家开展为期 300 天的鱼类疫病数据采集工作,并形成细粒度鱼类疫病识别数据集 FishData01。鱼类疫病图像主要在青岛市养殖基地晴天光照条件较好的情况下采集,由两位鱼类养殖专家审查数据集,以确认基本事实的有效性。由于采集的原始数据在训练时无法适应复杂的现场条件,在研究中为防止过拟合现象,提升训练模型的泛化能力,通过鱼类疫病图像数据增强方法提高训练样本的多样性,扩大数据集,增加所提模型的鲁棒性,从而满足模型所需的最大数据量。对训练收集到的图片进行图像旋转、锐度值操作以及对原始的细粒度鱼类疾病图像进行高斯处理,处理前后图像对比如图 7 所示。

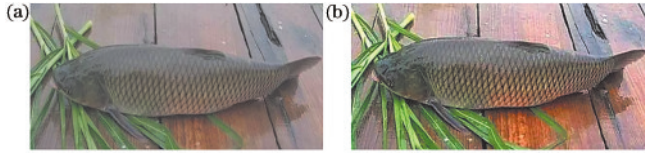


图 7 处理前后对比。(a)图像处理前数据;(b)图像处理数据
Fig. 7 Comparison before and after treatment. (a) Data before image processing; (b) data after image processing

表 1 测试机器配置

Table 1 Test machine configuration

Machine name	CPU	RAM	GPU	Operating system
LAPTOP-USO0 EFMM	AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz	16 GB	NVIDIA GeForce RTX 3060 Laptop GPU 6144 MB	Windows 11

3.3 评价标准

根据得到的混淆矩阵给出鱼病疫病检测结果,分别计算每种鱼类疾病的分类精度、敏感性和特异性。准确性是评价重要指标之一,式(10)为准确率的直观解释,即正确预测数量占总数的比例, $\sum P$ 为正确预测的次数, $\sum Q$ 为预测的总次数。对于二元分类可以采用式(11)计算。

$$R_{\text{accuracy}} = \frac{\sum P}{\sum Q} \times 100\%, \quad (10)$$

$$R_{\text{accuracy}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{FP}} + N_{\text{FN}} + N_{\text{TN}}} \times 100\%, \quad (11)$$

$$R_{\text{precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}} \times 100\%, \quad (12)$$

$$R_{\text{recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}} \times 100\%, \quad (13)$$

$$S_{\text{F1}} = \frac{2 \times R_{\text{precision}} \times R_{\text{recall}}}{R_{\text{precision}} + R_{\text{recall}}}, \quad (14)$$

式中: N_{TP} 、 N_{FP} 、 N_{FN} 、 N_{TN} 分别为真正样本、假阳性样本、假阴性样本、真负样本的数量。

在使用的数据集是平衡的条件下,精度是一个可靠的度量。式(12)为鱼类分类精度的百分比。式(13)

在本研究中,对所有分类不一致的图像进行分析,剔除不能得到代表性的图像。排除包含伪影图像(33张)、光照差图像(47张)、特征不明显图像(72张)等共 201 张图像,共收集鱼类疫病图像数据 2020 张,其中,传染性鱼病图像 499 张、非寄生物引起的鱼病图像 510 张、侵袭性鱼病图像 507 张、健康鱼类图像 504 张。将数据集按 4:1 的比例随机分为训练集、测试集。

为了缓解类别不平衡的问题,尝试两种不同的方法:1)在训练过程中给每个类别分配不同的权重,其中权重与各自类别中的样本数量成反比;2)过采样或数据增强,对鱼类疫病图像数据进行对比度增强等操作,使每个类有相同的样本数量。经过测试,数据增强可以平衡数据库从而提高分类准确率,为后续实验提供高质量的数据集。

3.2 算法环境规范

本次训练基于 NVIDIA GeForce RTX 3060, 16 GBRAM, Windows 11 系统, Tensorflow 1.7 框架, Python 3.7 实现。测试机器配置具体如表 1 所示。这些网络使用“提前停止”进行 100 个 epoch 的训练,计算每个网络的宏观精度和加权精度。

为敏感度或者召回率百分比。式(14)为 F1 评分,该指标为精确率和召回率的综合评价指标。研究中综合考虑 F1 评分和准确性作为模型性能的主要评价指标。

3.4 实验对比分析

为了验证所提 CViT-FDRM 细粒度分类效果,采用组归一化方法,分别在公开细粒度图像数据集与搭建的鱼类疫病图像数据库 FishData01 中进行对比实验。学习速率最初设置为 1×10^{-4} ,采用衰减学习率机理,将学习率降到最低的 1×10^{-6} ,根据交叉熵损失方法(MSE)衡量损失量,图 8 显示 CViT-FDRM 细粒度分类模型在 FishData01 数据集上的实验过程与结果。从图中可以看出,随着训练轮次的增加,0~25 轮训练集与测试集的分类准确率上升较快,25~75 轮上升较慢,在 75 轮后较为稳定,训练集分类准确率可达 99.98%,测试集分类准确率可达 97.02%。损失值在 0~25 轮下降较快,在 75 轮之后趋于稳定,最终损失值接近 0。由此可知, CViT-FDRM 细粒度分类模型可实现细粒度鱼类疫病识别。

3.4.1 错误结果分析

为验证 CViT-FDRM 细粒度分类模型效果,在搭建的鱼类疫病图像数据库 FishData01 下进行实验,将总数据量为 2020 张的鱼类疫病数据按 4:1 划分训练集

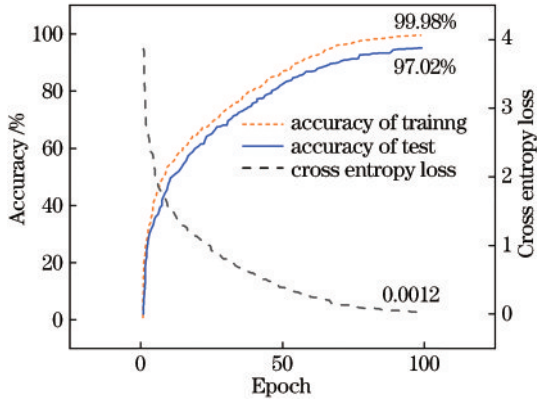


图8 CViT-FDRM 细粒度分类模型训练结果

Fig. 8 Training results of CViT-FDRM fine granular classification model

与测试集,总计4种鱼类数据。分类结果如表2所示,平均识别 $R_{accuracy}$ 为0.9702,平均识别 $R_{precision}$ 为0.9632,平

表2 CViT-FDRM 在 FishData01 数据集中分类效果

Table 2 Classification effect of CViT-FDRM in FishData01 dataset

Data size	Training set	Test set	Data category	$R_{accuracy}$	$R_{precision}$	R_{recall}	S_{F1}
2020	399	100	Infectious fish disease	0.9600	0.9505	0.9600	0.9552
	408	102	Non-parasitic fish disease	0.9902	0.9902	0.9902	0.9902
	406	101	Invasive fish disease	0.9703	0.9515	0.9703	0.9608
	403	101	Healthy fish	0.9604	0.9604	0.9604	0.9604
Average				0.9702	0.9632	0.9702	0.9667

3.4.2 不同归一化方法结果对比

CViT-FDRM 细粒度分类模型采用 GN 方式,在此分别与其他归一化方式:批归一化(BN)、LN、IN 进行对比实验,并进行可视化展示。BN 层取一批样本(batch 可设置为 2、4、6、8、16、32),对同一通道上的样本按批次进行归一化处理;LN 层是其中的一个样本在所有的通道上面进行归一化处理;IN 层是对某一样本在其中一个通道进行归一化处理;GN 层可以理解为 LN 层一个样本在一组通道上进行归一化处理,将所有通

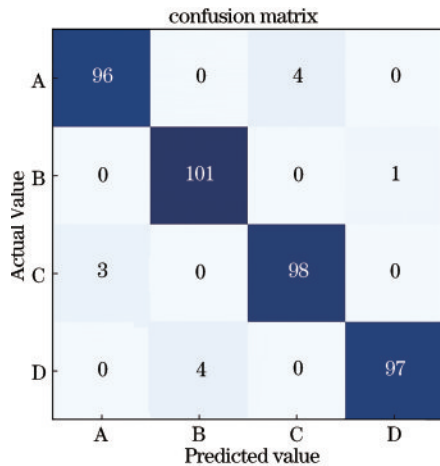


图9 CViT-FDRM 模型分类混淆矩阵

Fig. 9 Classification confusion matrix of CViT-FDRM model

均识别 R_{recall} 为 0.9702,平均识别 S_{F1} 为 0.9667,其中,非寄生生物引起的鱼病 $R_{accuracy}$ 最高为 0.9902。

为进一步确定模型的效果,采用混淆矩阵查看各类别分类效果,如图9所示,横坐标为预测值,纵坐标为真实值,A、B、C、D 分别代表传染性鱼病、非寄生引起的鱼病、侵袭性鱼病、健康鱼类。测试集总计 404 张鱼类疫病图像,图中深色对角线表示正确分类的鱼类数量。其中,正确分类的图像数据为 390(96+99+98+97)张,错误分类的图像为 14(404-390)张。对于 A 传染性鱼病,其中错误分类有 4 张(4 张预测为 C 侵袭性鱼病);对于 B 非寄生引起的鱼病,其中错误分类有 1 张(1 张预测为 D 健康鱼类);对于 C 侵袭性鱼病,其中错误分类有 3 张(3 张预测为 A 侵袭性鱼病);对于 D 健康鱼类,其中错误分类有 4 张(4 张预测为 B 非寄生引起的鱼病)。对未正确分类的图像进行进一步研究可知,主要原因各个类别图像相似度高,特征不明显。

道划分为不同组群,同样也可以理解为 IN 层将所有通道划分为通道数量的组群(通道数为组群数),划分的组群个数设置为 2、4、6、8、16、32。在 FishData01 数据集进行测试确定不同归一化方法的测试错误率并绘制 recall-precision(P-R)曲线直观表示分类精度。

图10(a)为 CViT-FDRM 模型采用 BN 方法时迭代轮次与分类误差图。可以看出:随着轮次的增加,误差损失不断减少;随着 batch 的增大,误差损失不断减少,batch 为 2 时,误差最小,但大 batch 对于深度学习来说输入数据量过大,不能满足要求。图10(b)为 CViT-FDRM 模型采用 GN 方法时迭代轮次与分类误差图。可以看出:随着轮次的增加,误差不断减少,相比于 BN 层来说,GN 层受输入批量 batch 大小影响较小,较为稳定。

为了更直观对比不同方法的效果,取 batch 为 32 绘制 P-R 曲线,结果如图11所示,曲线下方所围成的图形面积越大代表效果越好。可以直观看出,GN 方法可以实现更好的分类效果,因此合理选择 GN 层的组群个数可取得更好的分类结果。

3.4.3 不同分类模型结果对比

在公开细粒度数据集 Oxford Flowers(102 个类别,8189 张图片)上进行对比实验,按 4:1 划分数据集验证各个模型分类效果,结果如表3所示。其中,Params 表示参数量,Flops 表示浮点运算数。可以看

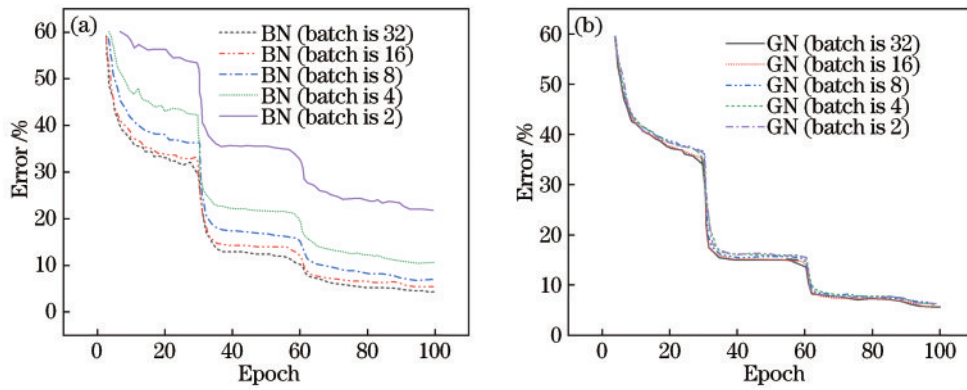


图 10 不同 batch 下 BN 与 GN 方法分类结果。(a) BN 分类结果图;(b) GN 分类结果图

Fig. 10 Classification results of BN and GN methods under different batches. (a) Classification results of BN; (b) classification results of GN

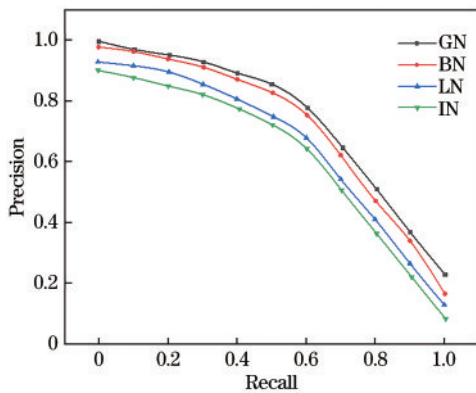


图 11 不同归一化方法 $P-R$ 曲线

Fig. 11 $P-R$ curves of different normalization methods

出,所提模型在开源细粒度数据集上相比于其他模型有较好的分类效果,其他 6 种模型平均识别精度为 0.9058, CViT-FDRM 的识别精度为 0.9542,相较于其他模型识别精度提高 4.84 个百分点。

表 3 经典模型分类效果对比

Table 3 Comparison of classification effects of classical models

Model	Accuracy	Params / 10^6	Flops / 10^9	Time /ms
MobileNetV2	0.8694	6.9	0.46	122
MobileNetV3	0.9103	4.1	0.22	60
EfficientNet-b2	0.9225	7.8	0.87	145
ShuffleNetV2	0.8541	5.6	0.79	136
ResNet18	0.9371	11.6	1.71	215
Vit-small	0.9414	21	4.20	380
CViT-FDRM	0.9542	8.9	1.29	159

4 结 论

合理结合 Transformer 和 CNN 结构的优势,改进 Transformer 结构,提出通用复杂环境下细粒度鱼类疫病识别模型 CViT-FDRM,解决鱼类疫病识别准确率、人工识别效率低等问题。1) 利用卷积结构对输入图像进行局部建模识别关键特征。CViT-FDRM 在

CNN 结构之后连接 Transformer 结构,使 CNN 结构聚焦于有用的特征,优化输出特征图,采用 GN 归一化方法合理划分通道,得到最终预测的重要特征,由此提高细粒度图像分类精度。2) 在复杂背景下建立的鱼类疫病数据集上的实验结果表明, CViT-FDRM 有较好的识别效果。在公开数据集上的识别结果明显领先于 MobileNetV3、Vit-small 与 EfficientNet-b2 等模型,表明所提模型是有效的疾病识别模型,具有实际应用价值。3) 构建细粒度鱼类疫病数据集 FishData01,采用图像数据增强方法提高训练样本的多样性与鲁棒性,可为相关研究提供研究支撑数据集。

参 考 文 献

- [1] Pauzi S N, Hassan M G, Yusoff N, et al. A review on image processing for fish disease detection[J]. Journal of Physics: Conference Series, 2021, 1997(1): 012042.
- [2] Park Y, Cho K H, Park J, et al. Development of early-warning protocol for predicting *chlorophyll-a* concentration using machine learning models in freshwater and estuarine reservoirs, Korea[J]. Science of the Total Environment, 2015, 502: 31-41.
- [3] Fang Y, Ramasamy R P. Current and prospective methods for plant disease detection[J]. Biosensors, 2015, 5(3): 537-561.
- [4] Arnal Barbedo J G. Digital image processing techniques for detecting, quantifying and classifying plant diseases [J]. SpringerPlus, 2013, 2: 660.
- [5] Sales J, Janssens G P J. Nutrient requirements of ornamental fish[J]. Aquatic Living Resources, 2003, 16 (6): 533-540.
- [6] Holm E A, Cohn R, Gao N, et al. Overview: computer vision and machine learning for microstructural characterization and analysis[J]. Metallurgical and Materials Transactions A, 2020, 51(12): 5985-5999.
- [7] Waleed A, Medhat H, Esmail M, et al. Automatic recognition of fish diseases in fish farms[C]//2019 14th International Conference on Computer Engineering and Systems (ICCES), December 17, 2019, Cairo, Egypt.

- New York: IEEE Press, 2019: 201-206.
- [8] Lin T Y, RoyChowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition[C]//2015 IEEE International Conference on Computer Vision, December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1449-1457.
- [9] Gao Y, Beijbom O, Zhang N, et al. Compact bilinear pooling[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 317-326.
- [10] Cui Y, Zhou F, Wang J, et al. Kernel pooling for convolutional neural networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3049-3058.
- [11] Kong S, Fowlkes C. Low-rank bilinear pooling for fine-grained classification[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 7025-7034.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12) [2022-09-05]. <https://arxiv.org/abs/1706.03762>.
- [13] Girdhar R, João Carreira J, Doersch C, et al. Video action transformer network[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 244-253.
- [14] 熊怡梦. 基于标签语义和 Transformer 的元学习小样本目标检测方法研究[D]. 西安: 西安电子科技大学, 2021: 1-83.
- Xiong Y M. Label semantics and Transformer for meta learning few-shot object detection[D]. Xi'an: Xidian University, 2021: 1-83.
- [15] Hatamizadeh A, Tang Y C, Nath V, et al. UNETR: transformers for 3D medical image segmentation[C]//2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), January 3-8, 2022, Waikoloa, HI, USA. New York: IEEE Press, 2022: 1748-1758.
- [16] Fu W H, McCalley J D, Vittal V. Risk assessment for transformer loading[J]. IEEE Transactions on Power Systems, 2001, 16(3): 346-353.
- [17] 何林远, 白俊强, 贺旭, 等. 基于稀疏 Transformer 的遥感旋转目标检测[J]. 激光与光电子学进展, 2022, 59(18): 1810003.
- He L Y, Bai J Q, He X, et al. Sparse Transformer based remote sensing rotated object detection[J]. Laser & Optoelectronics Progress, 2022, 59(18): 1810003.
- [18] 汪千金, 崔海华, 张益华, 等. 面向光学测量跨源点云的多尺度采样配准方法[J]. 光学学报, 2022, 42(10): 1015002.
- Wang Q J, Cui H H, Zhang Y H, et al. Multi-scale sampling registration method for optical measurement of cross-source point clouds[J]. Acta Optica Sinica, 2022, 42(10): 1015002.
- [19] 刘秀航, 黄宇辉, 张艳喜, 等. 基于 BP 神经网络补偿卡尔曼滤波的激光-MIG 复合焊缝熔宽在线检测[J]. 中国激光, 2022, 49(16): 1602011.
- Liu X H, Huang Y H, Zhang Y X, et al. Online weld width detection of laser-MIG hybrid welding based on Kalman filter algorithm compensated by BP neural network[J]. Chinese Journal of Lasers, 2022, 49(16): 1602011.
- [20] 胡杰, 刘汉, 徐文才, 等. 基于三维激光雷达的道路障碍物目标位姿检测算法[J]. 中国激光, 2021, 48(24): 2410001.
- Hu J, Liu H, Xu W C, et al. Position detection algorithm of road obstacles based on 3D LiDAR[J]. Chinese Journal of Lasers, 2021, 48(24): 2410001.
- [21] Meng L C, Li H D, Chen B C, et al. AdaViT: adaptive vision transformers for efficient image recognition[C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. New York: IEEE Press, 2022: 12299-12308.
- [22] 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.
- [23] Malik S, Kumar T, Sahoo A K. Image processing techniques for identification of fish disease[C]//2017 IEEE 2nd International Conference on Signal and Image Processing, August 4-6, 2017, Singapore. New York: IEEE Press, 2017: 55-59.
- [24] Chakravorty H, Paul R, Das P. Image processing technique to detect fish disease[J]. International Journal of Computer Science and Security (IJCSS), 2015, 9(2): 121-131.
- [25] Føre M, Frank K, Norton T, et al. Precision fish farming: a new framework to improve production in aquaculture[J]. Biosystems Engineering, 2018, 173: 176-193.