基于 FA-ConvNeXt 和小样本学习的唐卡主尊识别

白　科,史　伟,赵心怡,徐家明
(宁夏大学 信息工程学院,宁夏 银川 750021)

摘　要:针对唐卡主尊图像识别过程中,由于图像结构和纹理特征复杂、颜色绚丽且部分构图元素具有较高相似度而造成识别类别混淆的问题,提出了 FA-ConvNeXt 网络。首先,对于目前分类方法存在的数据集类别少、数量不平衡等问题,通过查阅资料和采用数据增强方法来扩充数据集。为了提高网络的分类准确度,在 ConvNeXt 网络架构上引入多尺度特征增强模块(MFEB),使网络更好地提取图像的结构和纹理特征,同时构建多注意力特征提取模块(MAEB),使网络更加关注具有判别性的特征,以减少冗余信息的干扰。通过实验与相关主流模型进行比较,结果表明,提出的 FA-ConvNeXt 网络识别准确率、召回率及 F1 值分别达到了 97.26%、97.18%、96.38%,较原网络分别提升了 7.35 百分点、6.94 百分点、6.17 百分点,且均优于被对比模型。最后将 FA-ConvNeXt 网络作为唐卡小样本学习的骨干网络,在小样本分类任务中也取得了良好的效果。

关键词:唐卡主尊图像识别;注意力;多尺度特征增强;小样本学习;ConvNeXt 网络
中图分类号:TP391-41;TN911.73-34　　**文献标识码**:A　　**文章编号**:1673-629X(2025)06-0027-07
doi:10.20165/j.cnki.ISSN1673-629X.2025.0022

Thangka Recognition via FA-ConvNeXt and Few-shot Learning

BAI Ke,SHI Wei,ZHAO Xin-yi,XU Jia-ming
(School of Information Engineering,Ningxia University,Yinchuan 750021,China)

Abstract:To address the issue of classification confusion in Thangka central deity image recognition due to complex textures,vibrant colors,and similar compositional elements,we introduce the FA-ConvNeXt network. Firstly,to tackle the problem of few categories and imbalanced quantities in current classification methods,the dataset is expanded through literature review and data augmentation techniques. To enhance the classification accuracy of the network,a Multi-scale Feature Enhancement Block (MFEB) is introduced on the ConvNeXt network architecture to better extract the structural and textural features of images. Additionally,Multi-Attention feature Extraction Block (MAEB) that integrates channel and multispectral channel attention is added to focus the network on discriminative features,thereby reducing the interference of redundant information. Experimental results compared with mainstream models show that the proposed FA-ConvNeXt network achieved recognition accuracy,recall,and F1 scores of 97.26%,97.18%,and 96.38%,respectively,which are 7.35 percentage points,6.94 percentage points,and 6.17 percentage points higher than the original network and superior to the compared models. Finally,the FA-ConvNeXt network is used as the backbone network for few-shot learning and has also achieved good results in few-shot classification tasks.

Key words:Thangka main deity recognition;attention;multi-scale feature enhancement;few-shot learning;ConvNeXt network

0　引　言

唐卡是藏族文化中的重要组成部分,更是中华民族弥足珍贵的非物质文化遗产。唐卡主尊图像的识别分类研究对于该非物质文化遗产的保护、推广和研究有着重要意义。王菽裕等[1]针对现有的唐卡头饰检测分类方法易受光照影响的问题,提出 AdaboostM2+HOG 算法,通过增强对图像特征的描述能力,提高了分类的准确度和鲁棒性。王铁君等[2]构造了多核支持向量机,实现了对唐卡图像和曼荼罗图像的有效分类,但两者都缺少对具体的唐卡主尊图像识别研究。

随着深度学习技术的快速发展更迭,基于神经网络的图像分类方法也相继被应用于唐卡主尊图像识别中。针对唐卡主尊图像复杂的特征,曾富亮等[3]提出了 L-DenseNet 网络,解决了 DenseNet 网络在特征传

收稿日期:2024-12-02　　**修回日期**:2025-03-06
基金项目:国家自然科学基金项目(62166030)
作者简介:白　科(1999-),男,硕士,研究方向为文物数字化保护和应用;通信作者:史　伟(1967-),女,教授,博士,研究方向为人工智能模式识别多学科交叉融合的文物古迹数字化保护应用。

播过程中存在丢失图像的负特征问题。在唐卡主尊分类任务中达到了 95.6% 的高准确率。陈玉红等[4]提出改进的基于卷积神经网络的唐卡尊像分类方法,在 19 895 张且含有 13 个类别的唐卡尊像数据集上达到了 94.7% 的准确率。薛盼盼等[5]提出的 SMADNet 网络,取得了 95.51% 的精度。然而,L-DenseNet 和 SMADNet 只是在类间相似度较小样本中进行训练,没有在类间相识度大的唐卡主尊进行进一步的研究。

为了解决这些问题,杨宇帆等[6]针对唐卡主尊身份的多粒度特征导致各细类别间类间距离不均衡的问题,提出了多层次分类器模型。将 ResNet-18 模型从原来 66.95% 的准确率提升至 71.47%。但人工构建的多层次关系图的准确性和完整性严重限制了网络的整体性能。

综上所述,唐卡主尊图像识别主要面临着三大挑战。第一,现有的网络不能很好地提取唐卡主尊图像复杂的结构和纹理信息,导致网络在类间相似度大、类内相似度小的样本图像中识别效果不佳;第二,缺少能够衡量模型整体性能的样本类别均衡的数据集;第三,部分模型尽管在类间相似度较大的样本中也取得了不错的分类精度,但模型在训练过程中需要大量的计算资源,使得计算过程耗时耗力,模型的性能甚至很大程度受限于人为因素。针对以上问题,该文通过查阅资料和图像增强方法,构建出类别均衡、质量较高的唐卡数据集。并在此基础上,提出 FA-ConvNeXt(Feature Augment ConvNeXt)网络模型。该网络引入多尺度特征增强模块,提取、增强图像的结构和纹理特征。通过添加融合多种注意力的特征提取模块,使得网络关注图像的判别性特征。改进后的网络可以充分发挥卷积网络和注意力网络的优势,实现模型整体的效能提升。

1 数据集和 FA-ConvNeXt 网络

1.1 数据集构建

类别均衡和高质量的数据集对于模型的训练至关重要。鉴于目前尚无公开的唐卡数据集,该文通过查阅资料和走访相关研究机构的方式进行数据收集,最终得到了 1 186 张高质量唐卡主尊图像。为了使训练出来的识别分类网络在类间相似度大和类内相似度小的样本上都具有良好的识别准确率,在对唐卡主尊数据集的主尊图像进行分类和统计后确定了七种主尊分类研究对象,分别是:宗巴喀大师、财宝天王(财神)、护法、佛母、莲花生大师、释迦摩尼、文殊菩萨。

通过随机旋转、颜色扰动和添加噪声等数据增强方法,使得图像数量扩充至 9 505 张。最终构建了 7 个主尊类别,数量为 9 505 张的唐卡主尊分类数据集,见表 1。

表 1 唐卡主尊数据集分布

唐卡类别	原始数据集	扩充后数据集
宗巴喀	185	1 480
财神	157	1 260
佛母	206	1 654
护法	144	1 152
莲花生	141	1 132
释迦摩尼	145	1 163
文殊菩萨	208	1 664

1.2 FA-ConvNeXt 网络结构

ConvNeXt 模型由 Facebook AI Research 等[7]提出。该文选取 T-ConvNeXt 作为基准模型,在网络中引入多尺度特征增强模块,并将多种注意力机制融合。构建了 FA-ConvNeXt 网络模型,如图 1 所示。图中,dim 为通道维度,该网络模型由多尺度特征增强模块(MFEB)和多注意力特征提取模块(MAEB)组成,旨在提取、增强唐卡图像的多尺度特征和关注图像的判别性特征。网络的机制如下:首先将尺寸大小为 224×224 的唐卡图像输入网络,经过第一个下采样模块,即大小为 4 × 4 的卷积层扩展该图像的通道数;然后利用 MFEB 模块提取、增强图像的多尺度特征并进行特征融合。再结合 MAEB 模块关注图像的判别性区域特征,消除冗余信息,以准确识别分类输入的唐卡主尊图像。

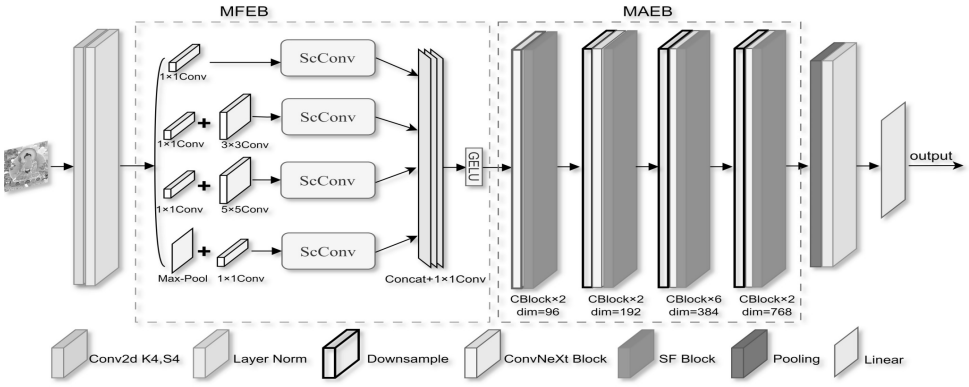


图 1 FA-ConvNeXt 网络结构

1.3　多尺度特征增强模块

针对唐卡主尊图像复杂的纹理和结构特征,设计多尺度特征增强模块(MFEB)。该模块对 GoogleNet 网络结构[8]的 Inception 块进行优化。设置不同大小的卷积核组合实现不同尺度的感知,然后在每个分支添加空间和通道重构卷积模块(ScConv)[9],减少冗余计算并促进代表性特征的学习,探究特征图上不同区域的相关性。最后将各分支进行通道融合,得到图像更好的表征。

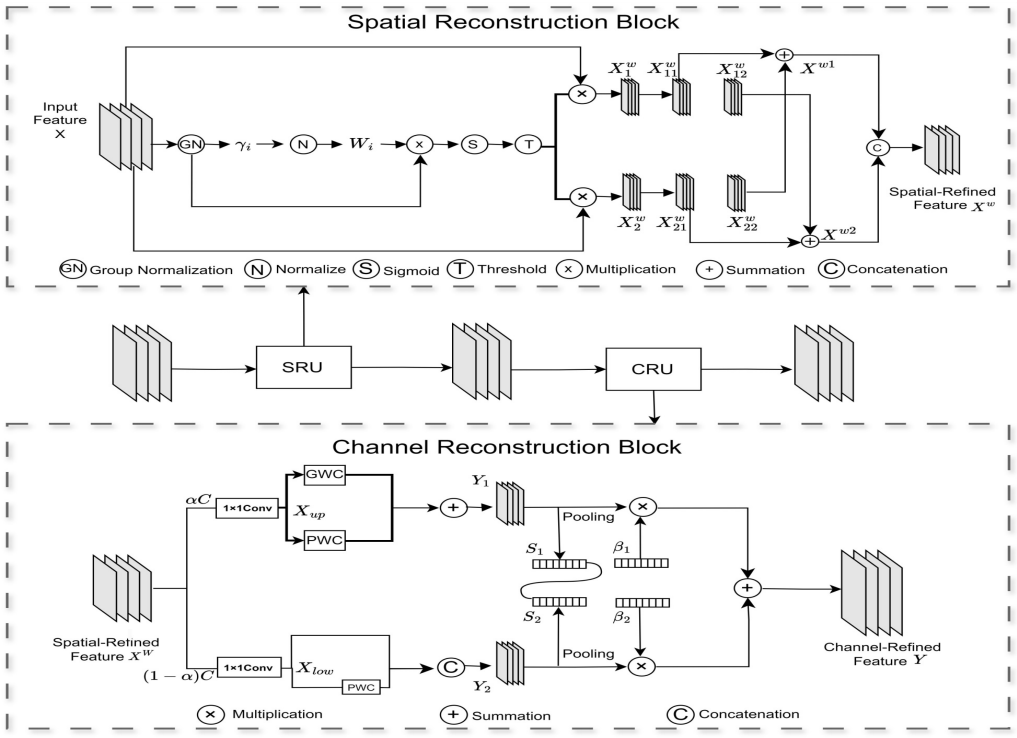


图 2　ScConv 网络结构

ScConv 由空间重构单元(SRU)和信道重构单元(CRU)组成,如图 2 所示。输入特征首先通过 SRU 里的一系列操作,获得空间细化特征 X^W,然后利用 CRU 运算获得信道细化特征。当给定一个特征映射 $X \in \mathbb{R}^{N\times C\times H\times W}$,减去平均值 μ 并除以标准差 σ 来标准化输入特征 X,如下所示:

$$X_{\text{out}} = \text{GN}(X) = \gamma \frac{X-\mu}{\sqrt{\sigma^2+\varepsilon}} + \beta \tag{1}$$

其中,μ 和 σ 是 X 的均值和标准差,ε 是为了避免分母为零时造成异常而加入的一个小的正常数,γ 和 β 是可训练参数。GN 层中的可训练参数 $\gamma \in R^C$ 用于计算每个批次中各通道的空间像素方差。归一化相关权重 $W_\gamma \in R^C$ 由公式 2 得到,表示不同特征映射的重要性。

$$W_\gamma = \{w_i\} = \frac{\gamma_i}{\sum_{j=1}^{C}\gamma_j}, \quad i,j = 1,2,\cdots,C \tag{2}$$

然后将经 W_γ 重新加权的特征映射的权值通过 sigmoid 函数映射到(0,1)范围,并通过阈值进行门控。这里将阈值以上的权重设置为 1,得到信息权重 w_1,将其设置为 0,得到非信息权重 W(实验中阈值设置为 0.5)。获取 W 的整个过程可以用公式 3 表示。

$$W = \text{Gate}(\text{Sigmoid}(W_\gamma(\text{GN}(X)))) \tag{3}$$

最后将输入特征 X 分别乘以 W_1 和 W_2,得到两个加权特征:信息量较大的特征 X_1^ω 和信息量较小的特征 X_2^ω。这样就成功地将输入特征分为两部分:X_1^ω 具有信息量和表达性的空间内容,而 X_2^ω 只含有较少的信息。重构操作中将信息丰富的特征与信息较少的特征相加,生成信息更丰富的特征,从而节省空间。采用交叉重构运算,将加权后的两个不同的信息特征充分结合起来,加强它们之间的信息流。然后将交叉重构的特征 X^{ω_1} 和 X^{ω_2} 在通道维度上进行拼接(Concatenation)操作,得到空间精细特征映射 X^ω。过程表示如下:

$$\begin{cases} X_1^\omega = W_1 \otimes X \\ X_2^\omega = W_2 \otimes X \\ X_{11}^\omega \oplus X_{22}^\omega = X^{\omega_1} \\ X_{21}^\omega \oplus X_{12}^\omega = X^{\omega_2} \\ X^{\omega_1} \cup X^{\omega_2} = X^\omega \end{cases} \tag{4}$$

其中,⊗是逐元素的乘法,⊕是逐元素的求和,∪ 是 Concatenation 操作。将 SRU 应用于中间输入特征 X 后,将信息较少的特征分离,再次重构,增强其判别性特征,抑制空间维度上的冗余特征。然而,空间精细特征映射 X^ω 在通道维度上仍然是冗余的。

最后在 CRU 单元中,将输入的空间细化特征 X^ω

分割成两个部分,一部分为通道数 αC 的特征,另一部分为通道数 $(1-\alpha)C$ 的特征,随后对于两组特征的所有通道利用 1×1 卷积压缩处理,分别得到 X_{up} 和 X_{low}。在转换操作中,将输入的 X_{up} 作为“富特征提取”的输入,分别进行 GWC(Groupwise Convolutional Filter)和 PWC(Pointwise Convolutional Filter),然后相加得到输出 Y_1,将输入 X_{low} 作为“富特征提取”的补充,进行 PWC,得到的记过和原来的输入取并集得到 Y_2。在融合操作中,采用了一种简化的 SKNet 来实现对特征图 Y_1 和 Y_2 的融合。首先,通过全局平均池化操作,将其中的空间信息与通道统计信息进行整合,得到池化后的特征表示 S_1 和 S_2。对其实施 Softmax 函数操作,以此生成特征权重向量 β_1 和 β_2。最终,依据这些特征权重向量,以产生最终的输出 $Y=\beta_1Y_1+\beta_2Y_2$,Y 即为通道提炼的特征。

1.4 多注意力特征提取模块

由于唐卡主尊图像的颜色和纹理结构复杂,因此,添加 SENet(Squeeze-and-Excitatio Attention)[10]学习图像各个通道的依赖关系,适应地调整对不同通道的关注度,更好地处理复杂颜色特征。FcaNet(Frequency Channel Attention)[11]利用了离散余弦变换(DCT)的多个频率成分,这样可以引入更多的信息解决单一的通道注意力中信息不充分的问题。由此,构建了多注意力特征提取模块(MAEB),该模块充分结合了 SENet 和 FcaNet 注意力的优点,如图 3 所示。MAEB 由多个 Downsample Block 和 ConvNeXt Block 堆叠而成。在对 MAEB 结构进行调整时,将原来的 ConvNeXt 模型中 ConvNeXt Block 的堆叠方式由(3,3,9,3)变为(2,2,6,2)。实验验证表明,这样的调整在对模型整体性能影响不大的情况下,可以减少模型的参数数量,降低计算复杂度。然后在每组 ConvNeXt Block 后额外增加一个由通道注意力和多谱通道注意力机制构成的混合注意力块(SF Block)。

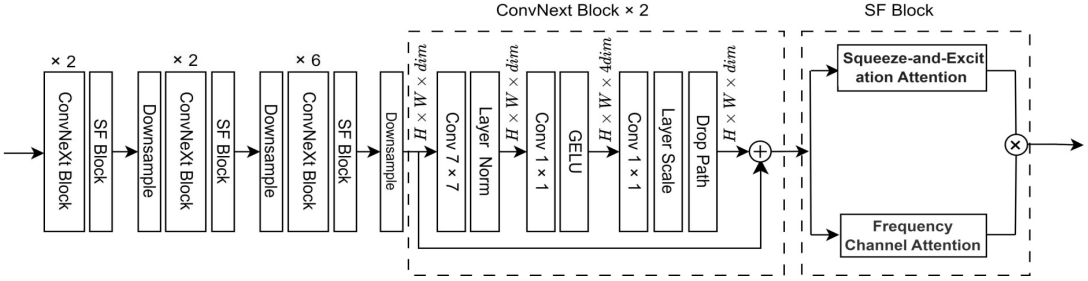


图 3 MAEB 模块结构示意图

FcaNet 可以学习和强调更重要的频率成分,更有效地聚焦于对识别任务有益的频率特征。SENet 更关注唐卡图像判别性特征并抑制不重要的特征,将两者提取的特征在通道的维度上相乘融合,能够更有效地捕捉到图像重要区域的细节和纹理信息。综上,这样的模块设计在不降低网络性能的同时,又提升了模型对图像的理解能力,减少了网络的参数量,进而缩短模型训练的时间。

2 基于 FA-ConvNeXt 骨干网络的小样本识别

由于部分唐卡主尊图像类别繁多,而数量极少,依赖大规模数据集的网络模型难以适应于小样本分类任务。因此,该文构建了 20 个类别的唐卡主尊小样本数据集,并设计一种基于 FA-ConvNeXt 骨干网络的小样本图像分类算法,如图 4 所示。

该算法采用基于原型网络[12]的小样本学习方法,将 FA-ConvNeXt 作为原型网络分类器的骨干网络。骨干网络分别对支持集和查询集中的图像进行特征提取,得到类别原型和嵌入编码。原型网络通过计算每个类别原型和查询集嵌入编码之间的相似度,以实现对图像的识别和分类任务。

算法设计:

首先,小样本学习需要构建多个小任务,每个任务中有 N 个类别,每个类别有 K 张图片,称为 N-way K-shot 图像分类任务,一共有 episodes 个小任务,每个小任务由以下两部分组成:

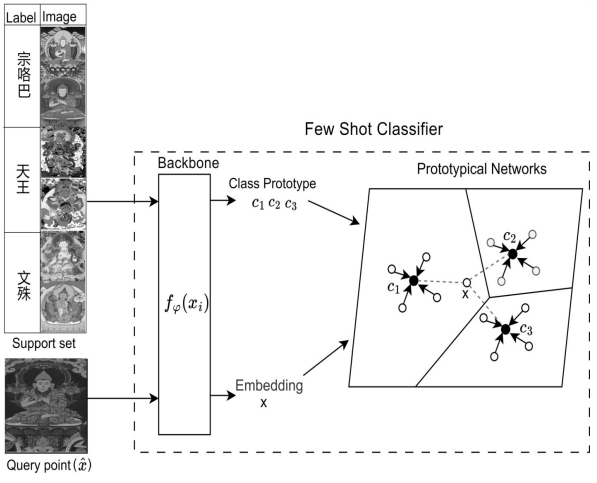


图 4 少样本分类器

(1)支持集(Support Set):相当于每个小任务中的训练集,包含 N 个分类标签,每个标签有 K 张图片。

(2)查询集(Query Set):相当于每个小任务中的测试集,包含 Q 张未分类的图片。

在唐卡主尊小样本数据集 D 中随机选取 N 个类别,每个类别选取 K 张图片,构成支持集,选取 Q 张图片,构成查询集,这样就组成一个 episode 的小数据集。以此类推,构造一个 episodes 小数据集。即:

$$D_{\text{episode}} = D_{\text{support}} \cup D_{\text{query}} = \{s_i\}_{i=1}^{n_s} \cup \{q_i\}_{i=1}^{n_q} \tag{5}$$

支持集中的每张图片利用骨干网络 $f_\varphi(x_i)$ 进行信息提取。

$$\begin{cases} hs_i = f_\varphi(s_i) \\ hq_i = f_\varphi(q_i) \end{cases} \tag{6}$$

学习到每张图片的 Embedding 编码表示。然后对支持集的每个类别下的 Embeddings 做均值处理,计算得到每个类别的原型表示(class Prototype)。其中,l_{s_i} 表示图片 s_i 的类别标签。

$$p_{c_j} = \sum_{\{i \mid l_{s_i} = c_j\}} hs_i \tag{7}$$

接着使用骨干网络将查询集图片进行编码,得到该图片的 Embedding 向量表示。并与类别原型进行欧氏距离的相似度计算。

$$p(\hat{l}_{q_i} = c_j) = \frac{\exp(\text{sim}(hq_i, p_{c_j}))}{\sum_{k=1}^{|c|} \exp(\text{sim}(hq_i, p_{c_k}))} \tag{8}$$

使用 softmax 运算将相似度激活成概率分布。将得到的查询集图片的标签和真实标签做交叉熵 loss,作为损失函数,然后梯度反向传播即可完成一个 episode 的训练。

$$L_{q_i} = \text{CrossEntropy}(l_{q_i}, \hat{l}_{q_i}) \tag{9}$$

最终将训练好的小样本分类器在测试集上进行测试。

3　结果与分析

3.1　实验环境及参数设置

为了综合衡量 FA-ConvNeXt 网络的性能,基于唐卡主尊数据集设计了消融和对比实验。在 Windows 平台下,使用 pytorch 框架进行模型的实现和训练,版本为 CUDA 12.5、pytorch 2.3.0,GPU 为 NVIDIA RTX A6000,显存大小为 48G。所有输入图片的大小为 224 × 224,设置批处理大小为 32,迭代次数为 50。

3.2　性能评价指标

选用准确率(ACC)、召回率(Recall)、F1 值(F1-Score)、混淆矩阵(Confusion Matrix)、帧率(FPS)、模型参数量 P(ParaFA)和推理时间(InferenceTime)作为所有网络模型的评价指标。

(1)准确率(Accuracy)。

准确率(Accuracy)通过计算 TP(真正例)、FP(误正例)、TN(真反例)和 FN(误反例)四个基础指标得出,其计算公式如式 10 所示。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{10}$$

(2)帧率(FPS)。

公式 11 表示帧率(Frames Per Second,FPS)的计算公式,其中 frameCount 表示检测图片的总数,elapsedTime 表示耗费的总时间。

$$\text{FPS} = \frac{\text{frameCount}}{\text{elapsedTime}} \tag{11}$$

(3)网络参数量(Params)。

网络参数量(Params)用来衡量网络模型大小,单个卷积核的参数量的计算如公式 12 所示。

$$\text{Params} = (K \times K \times C_{\text{in}}) \times C_{\text{out}} + C_{\text{out}} \tag{12}$$

其中,C_{in} 表示卷积层输入张量的通道数,C_{out} 表示卷积层输出张量的通道数,K 表示卷积核的大小。

(4)推理时间(InferenceTime)。

公式 13 中,EndTime 和 StartyTime 分别表示模型推理过程的结束时间和开始时间。

$$\text{InferenceTime} = \text{EndTime} - \text{StartTime} \tag{13}$$

(5)召回率(Recall)。

召回率是指分类正确的正样本个数(TP)占真正的正样本个数(TP+FN)的比例。召回率又称查全率,具体的公式如下:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{14}$$

(6)F1 值(F1-Score)。

精确率和召回率的调和平均值,计算公式如下:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{15}$$

3.3　消融实验

为了验证 FA-ConvNeXt 的有效性,设置了 4 组消融实验。主干网络选用 ConvNeXt 网络,结果如表 2 所示。

表 2 中,"√"表示采用了改进算法。例如:MFEB-ConvNeXt 表示采用了 MFEB 的 ConvNeXt。分析表中数据可以看出,MFEB-ConvNeXt 较原网络在准确率上提高 4.82 百分点,验证了加入 MFEB 模块能够提高网络模型对复杂唐卡图像的识别效果。MAEB-ConvNeXt 网络的分类准确率达到了 93.99%,证实了改进的 MAEB 模块能够有效提升模型的效能。纵向对比 4 个网络的实验数据可知,FA-ConvNeXt 简洁的网络设计和高效的性能,使其在唐卡识别研究中更具优势。

表 2 消融实验结果

网络模型	FPS/(帧·s^{-1})	T/FA	改进算法		ACC/%	F1/%	*R*/%
			MFEB	MAEB			
ConvNeXt	103.4	9.1	—	—	89.91	90.21	90.24
MFEB-ConvNeXt	84.2	11.2	√	—	94.73	94.88	94.95
MAEB-ConvNeXt	99.99	9.5	—	√	93.99	93.94	93.98
FA-ConvNeXt	88.8	10.6	√	√	97.26	96.38	97.18

3.4 与原模型对比实验

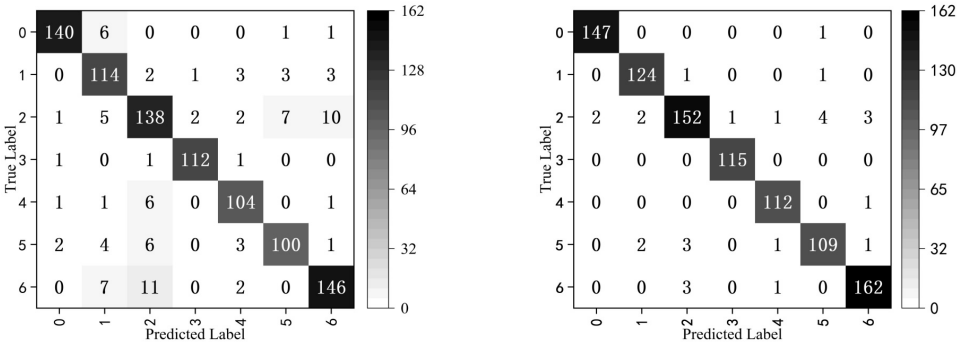
改进前、后网络的混淆矩阵如图 5 所示,虽然 ConvNeXt 网络大多数分类结果集中在混淆矩阵的对角线上,但在对类别 1(财神)、类别 2(佛母)和类别 4(莲花生大师)的区分上仍存在混淆情况,这说明 ConvNeXt 网络对于类间相似度大的样本识别能力不足。而 FA-ConvNeXt 网络的分类结果更加集中地落在混淆矩阵的对角线上,尤其在类别 3(护法)的识别结果中,分类准确率达到 100%。

结合表 3、表 4 及图 5 可知,相较于改进前的 ConvNeXt 网络,改进后的网络对于类间相似度大和类内相似度小的唐卡主尊图像都能够较好地识别和区分。此外,识别能力较原网络更强,网络模型也更加轻量。

3.5 与其他模型对比实验

通过与目前图像识别性能较好的网络模型进行对比,得出具体的评价指标,以客观准确地评价 FA-ConvNeXt 的模型性能。实验结果如表 3 所示。

可以看出 FA-ConvNeXt 网络相较于 ResNet-152[13]、Vision-LSTM[14]、Swin Transformer[15]、ConvNeXt-V2[16]、Mobile VIT[17]、L-DenseNet[3] 网络,其准确率、F1 值和召回率均有所提升。Mobile VIT 通过结合 CNN 和 ViT 的优势,在识别效率和模型轻量化方面表现较好。但 FA-ConvNeXt 主要应用于服务端环境,其对于模型轻量化的实际需求相对较低,故在识别准确度上的优势使其更加适合对准确性要求高的研究场景。



(a) ConvNeXt (b) FA-ConvNeXt

图 5 改进前、后网络的混淆矩阵对比

表 3 不同模型在测试集上的比较

网络模型	FPS/(帧·s^{-1})	T/FA	P/10^6	ACC/%	F1/%	*R*/%
ResNet152	76.3	12.4	58.15	90.17	90.64	90.94
Vision-LSTM	39.5	23.9	6.08	79.83	80.14	81.88
Swin Transformer	78.5	12.1	27.52	93.25	93.18	93.37
ConvNeXt v2	66.19	14.3	88.57	87.14	87.11	87.52
Mobile VIT	87.8	10.7	0.95	95.04	95.25	95.14
L-DenseNet	75.7	12.5	7.97	93.25	93.14	93.28
ConvNeXt	103.4	9.1	27.81	89.91	90.21	90.24
FA-ConvNeXt	88.8	10.6	19.47	97.26	96.38	97.18

3.6 小样本分类实验

为进一步衡量 FA-ConvNeXt 网络对唐卡主尊图像的特征提取能力,基于唐卡主尊小样本数据集设计了小样本分类实验。使用 pytorch 框架,版本为 CUDA

11.8、pytorch 2.2.2,GPU 为 NVIDIA GeForce 3080ti,输入图片的大小为 224×224,具体如下:

在小样本数据集中,分别由 10 个类别组成训练集,5 个类别组成验证集,5 个类别组成测试集。设置批处理大小为 32,迭代次数为 50,在训练集上对骨干网络进行训练。每训练 10 个 epoch,对少样本分类器进行验证,以此得到效果最佳的模型。

最后,利用测试集对基于 FA-ConvNeXt 骨干网络的少样本分类器进行小样本分类任务测试,设置测试任务数为 100,query=2,采用 3-way,1-shot 和 3-way,3-shot 的形式来评估模型在不同难度级别下的泛化能力。以所有测试任务的平均分类准确率作为本实验的性能指标。实验结果见表 4。

表 4　在小样本数据集上的准确率对比　%

骨架网络	3-way 1-shot	3-way 3-shot
ConvNeXt	67.97	74.28
MFEB-ConvNeXt	68.61	76.16
MAEB-ConvNeXt	69.35	77.59
FA-ConvNeXt	72.22	83.33

实验证明,基于 FA-ConvNeXt 骨干网络的分类器在 3-way 1-shot 和 3-way 3-shot 的小样本识别分类任务中,识别准确率分别达到了 72.22% 和 83.33%,均优于基于 ConvNeXt 骨架网络的分类器,有效说明了 FA-ConvNeXt 网络设计的合理性,在提取图像的特征能力方面,具有显著的优势,也表明了该文设计的小样本分类器性能良好。

4　结束语

研究了唐卡识别分类问题,在 ConvNeXt 网络基础上,通过引入多尺度特征增强模块、添加融合多种注意力的特征提取模块的操作,最终提出了 FA-ConvNeXt 网络。然后对唐卡主尊数据集的图像进行分类,结果表明,FA-ConvNeXt 网络的识别准确率、F1 分数和召回率均得到了提升,并且其性能要优于相关主流模型。最后在小样本分类实验中,进一步证明了 FA-ConvNeXt 网络相较于 ConvNeXt 网络具有更好的特征提取能力。

参考文献:

[1] 王菽裕,宋俊芳,张春玉. Adaboost M2+HOG 算法在肖像类唐卡图像头饰检测分类中的应用[J]. 无线互联科技,2023,20(11):99-102.

[2] WANG T,WANG W. Research on a Thangka image classification method based on support vector machine[J]. International Journal of Pattern Recognition and Artificial Intelligence,2019,33(2):14-23.

[3] 曾富亮,胡文瑾,何国源,等. 基于 DenseNet 的唐卡图像分类[J]. 现代电子技术,2022,45(6):153-157.

[4] 陈玉红,刘晓静. 基于卷积神经网络的唐卡尊像自动分类研究[J]. 计算机技术与发展,2021,31(12):167-174.

[5] 薛盼盼. 基于深度学习的唐卡主尊分类及小样本目标检测算法研究[D]. 兰州:西北民族大学,2022.

[6] 杨宇帆. 唐卡图像主尊多层次特征学习与分类研究[D]. 拉萨:西藏大学,2023.

[7] LIU Z,MAO H,WU C Y,et al. A convnet for the 2020s[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New Orleans: IEEE/CVF,2022:11976-11986.

[8] SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Boston:IEEE,2015:1-9.

[9] LI J,WEN Y,HE L. Scconv:spatial and channel reconstruction convolution for feature redundancy[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. Vancouver:IEEE/CVF,2023:6153-6162.

[10] HU J,SHEN L,SUN G. Squeeze-and-excitation networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Salt Lake City: IEEE,2018:7132-7141.

[11] QIN Z,ZHANG P,WU F,et al. Fcanet:frequency channel attention networks[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal: IEEE/CVF,2021:783-792.

[12] SNELL J,SWERSKY K,ZEMEL R. Prototypical networks for few-shot learning[C]//Proceedings of the 31st international conference on neural information processing systems. Red Hook:Curran Associates Inc.,2017:4080-4090.

[13] HE K,ZHANG X,REN S,et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. Las Vegas: IEEE,2016:770-778.

[14] ALKIN B,BECK M,PÖPPEL K,et al. Vision-LSTM:xLSTM as generic vision backbone[J]. arXiv:2406.04303,2024.

[15] LIU Z,LIN Y,CAO Y,et al. Swin transformer:hierarchical vision transformer using shifted windows[C]//Proceedings of the IEEE/CVF international conference on computer vision. Montreal:IEEE/CVF,2021:10012-10022.

[16] WOO S,DEBNATH S,HU R,et al. ConvNeXt v2:co-designing and scaling convnets with masked autoencoders[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR '23). Vancouver:IEEE/CVF,2023:16133-16142.

[17] MEHTA S,RASTEGARI M. Mobilevit:light-weight,general-purpose,and mobile-friendly vision transformer[J]. arXiv:2110.02178,2021.