

基于改进 ResNet 深度学习的古代壁画分类方法*

曹建芳^{1,2} 彭存赫¹ 陈志强³ 杨卓林¹

(1. 太原科技大学计算机科学与技术学院 太原 030024; 2. 忻州师范学院计算机系 忻州 034000;

3. 山西工程技术学院大数据与智能工程系 阳泉 030002)

摘要: 针对壁画图像人物间纹理,轮廓相似,不同场景下壁画人物特征差异较大,背景噪声复杂,分类易混淆等问题,提出了一种针对 ResNet 卷积神经网络的改进策略。首先将模型输入层中较大的 7×7 卷积核分离为 3 个串联的 3×3 小卷积核堆积的主干,将 2×2 平均池化与最大池化进行 add 特征融合取代原最大池化操作,增强模型的表征能力。其次设计了一种多尺度高效的通道注意力模块,以 ECA 通道注意力模块为基础,串联空间注意力模块,将空间模块中原 3×3 卷积核替换为 SK 注意力模块,融合多尺度信息捕捉全局长距离依赖关系,降低背景噪声的干扰。最后提出一种蜂窝式聚合结构,将相邻的 block 块中的输出信息进行 add 操作,作为后续层的输入,同时捕获低级和高级特征,增强上下文信息的流通性。实验结果表明:该模型在准确率、精度、召回率和 F1 值分别达到 96.51%、96.65%、96.67%、96.63%。相对于原模型 ResNet-18 准确率提升 9.76%,与主流的分类算法相比分类准确率、泛化能力、稳定性均有一定的提升,能够高效准确识别壁画所属类型,这对于文化遗产保护和艺术史方面研究具有显著价值。

关键词: 壁画分类; ResNet; 注意力机制; 特征提取; 卷积神经网络; 深度学习

中图分类号: TP391; TN03 **文献标识码:** A **国家标准学科分类代码:** 510.1050

Classification of ancient murals based on improved ResNet deep learning

Cao Jianfang^{1,2} Peng Cunhe¹ Chen Zhiqiang³ Yang Zhuolin¹

(1. School of Computer Science and Technology, Taiyuan University of Science and Technology, Taiyuan 030024, China;

2. Department of Computer, Xinzhou Normal University, Xinzhou 034000, China;

3. Department of Big Data and Intelligent Engineering, Shanxi Institute of Technology, Yangquan 030002, China)

Abstract: Aiming at texture problems, contour similarity among fresco image characters, large differences in fresco character features in different scenes, complex background noise, and confusing classification, an improvement strategy for ResNet convolutional neural network is proposed. Firstly, the larger 7×7 convolutional kernel in the input layer of the model is separated into three series-connected 3×3 small convolutional kernels stacked in the backbone, and 2×2 average pooling and maximum pooling are used for add feature fusion to replace the original maximum pooling operation, which enhances the model's representative ability. Secondly, a multi-scale efficient spatial channel attention module is designed, based on the ECA channel attention module, the spatial attention module is connected in series, and the original 3×3 convolutional kernel in the spatial module is replaced by the SK attention module, which fuses the multi-scale information to capture the global long-distance dependency, and reduces the interference of background noise. Finally, a cellular aggregation structure is proposed to perform ADD operation on the output information in the neighboring block blocks as inputs to the subsequent layers, capturing both low-level and high-level features to enhance the circulation of contextual information. The experimental results show that the model achieves 96.51%, 96.65%, 96.67% and 96.63% in accuracy, precision, recall and F1 value, respectively. Relative to the original model ResNet-18 accuracy is improved by 9.76%, and compared with mainstream classification algorithms classification accuracy, generalization ability, and stability are all improved, which can efficiently and accurately identify the type of mural belonging to the mural, which is of significant value for cultural heritage preservation and art history aspects of the research.

Keywords: mural classification; ResNet; attention mechanism; feature extraction; convolutional neural network; deep learning

0 引言

壁画作为我国重要的文化遗产,承载着古代文明的科

技与辉煌,体现着当时社会先进生产力,具有很高的研究价值^[1]。其中,壁画分类又是研究和保护古代壁画的重要一环。因此本文针对古代壁画旨在研究一种能够快、准确实

收稿日期:2024-06-30

* 基金项目:国家自然科学基金面上项目(62372397)、教育部人文社会科学研究项目(规划基金项目)(21YJAZH002)、山西省自然科学基金面上项目(202203021221222)、山西省文物局 2024 年度文物科研课题(2024KT23)项目资助

现壁画精准分类的高效算法。从而提升古代壁画数字化保护与研究的自动化水平,促进艺术与技术的跨学科融合。

传统的古代壁画图像分类研究主要通过手工设计的特征提取器提取壁画特征,再使用机器学习算法进行分类。郝亚博等^[2]提出通过计算图像之间的相似程度进行壁画分类。钱文华等^[3]提出通过求取图像的颜色熵、分块熵、轮廓熵,合并构成不同风格的信息熵,通过支持向量机对壁画图像学习训练。Zeng 等^[4]提出通过提取壁画特征,生成视觉字典,并结合色调饱和和度值(hue saturation value, HSV)提取的颜色特征进行壁画分类。Zhang 等^[5]提出使用光谱匹配算法(spectral matching algorithm, SMA)结合激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)的方法对莫高窟壁画颜料进行识别。上述研究主要通过机器学习算法提取壁画的颜色、形状、纹理等特征,进行壁画分类。特征提取过程繁琐单一,往往局限于表面特征的描述。假如能够深度挖掘更细微的特征描述,加大深层次文化内涵和意义的解读,识别精度仍有提升空间。

随着深度学习的发展,卷积神经网络由于其在计算机视觉领域具有良好的分类性能而得到了广泛关注和应用^[6]。2012 年图像分类竞赛 Krizhevsky 等^[7]提出使用深层 CNN 结构网络获取了冠军。之后随着 VGGNet^[8]、GoogleNet^[9],再到 ResNet^[10]、DenseNet^[11]、EfficientNet^[12-13]的出现,利用卷积神经网络进行图像分类已经取得显著效果。李丹等^[14]提出使用轻量卷积网络进行钢表面缺陷分类。张晋婧等^[15]提出在 FaceNet 网络基础上融合注意力机制提取人脸特征。然而,目前应用 CNN 对壁画图像进行分类的研究仍处于起步阶段。Li 等^[16]提出使用复杂的颜色形状描述符结合深层卷积网络来描述绘图样式对壁画进行朝代分类,二元分类达到 91.62%。而王燕妮等^[17]提出从壁画颜料分类角度出发,用多光谱图像的分类问题转化为解决壁画颜料的分类问题,准确率达到 89.18%。之后 Cao 等^[18]提出采用小卷积核提取壁画浅层背景特征,之后将较大的卷积核十字分解为更小的卷积核提取重要的深层特征信息进行壁画分类,准确率提升至 91.38%。以上方法多数是通过传统方式提取壁画的外观特征,再结合深度学习思想进行壁画分类。虽然性能有所提升,但无法跨越“语义鸿沟”,缺乏深层语义信息的提取能力。

综上,目前图像分类算法存在特征提取过程繁琐单一,局限于表面特征的描述;以及对于复杂图像无法跨越“语义鸿沟”问题。因此以壁画图像作为研究对象需要同时考虑如下问题:1)古代壁画不同人物之间特征相似,难以区分;2)背景噪声的干扰;3)不同场景下壁画人物特征差异较大。对此本文以提出了一种基于改进 ResNet 深度学习的古代壁画分类方法。预期增强人物特征区分能力,有效抑制背景噪声的干扰,适应不同场景下的人物特征变化。主要贡献如下:

1)提出一种浅层特征深度提取模块(shallow feature depth extraction model, SD-Model),将模型输入层中较大的 7×7 卷积核分离为小卷积核堆积的主干,将平均池化与最大池化进行 add 特征融合取代原最大池化操作,并将所有步长取值为 1,生成更精细的特征表示,避免信息丢失和过度压缩。

2)提出一种多尺度高效的空间通道注意模块(multi-scale efficient spatial channel attention, MESCA),以 ECA^[19]通道注意力模块为基础,串联空间注意力模块,将空间模块中原 3×3 卷积核替换为 SK 注意力模块^[20],通过融合多尺度信息捕捉全局长距离依赖关系,高效地获得全面而具有区分性的特征表示。

3)提出一种蜂窝式聚合结构(honeycomb polymerisation structure, HP-Structure),借鉴 DenseNet 网络密集连接的思想,将相邻的 block 块中的输出信息进行 add 操作,作为后续层的输入,用来综合壁画整体结构信息和细节信息,增强上下文信息的流通性,缓解“语义鸿沟”问题。

1 基本理论

1.1 ResNet-18 模型

ResNet-18 是一种深度卷积神经网络,结构如表 1 所示,主要由 4 个 block 块组成。理论上讲,随着神经网络的深入,其非线性表示的能力和识别效果都会得到提高。然而,由于深度的网络结构,在梯度反向传播的过程中会产生“连乘”现象,造成梯度爆炸和梯度消失等现象。ResNet-18 采用残差学习单元结构,从而解决了深层网络梯度消失问题。ResNet-18 使用的残差学习单元如图 1 所示,整个残差学习单元除了常规的通过权重层产生的输出外,还设计了一条分支,允许输入直接传递至输出,并与此权重层的输出进行结合。设输入为 x ,经过学习后的最终输出为 $H(x)$,权重层的独立输出表示为 $F(x)$,那么整个残差学习过程中可以表达为:

$$H(x) = F(x) + x \quad (1)$$

如果权重层没有学习到图像特征,即 $F(x)$ 为 0,则 $H(x)$ 为一个恒等映射(identity mapping)。把当前输出直接传递给下一层网络,在梯度反向传播过程中,下层网络的梯度同样直接传递给上一层网络,在不产生额外的计算量的情况下,直接解决了网络过深而导致的梯度消失问题。ResNet-18 深层次的卷积结构可以学习壁画图像抽象的深层语义信息,理解图像信息更多内在的关联;而残差学习单元可以更好地保持梯度的稳定性,也不会增加计算的复杂度,使其允许逐层深入地表征更多的特征。对于内容丰富,语义复杂的古代壁画图像,可以获得更高的识别性能。

1.2 注意力机制

受到人类视觉和认知过程的启发,注意力机制被引入卷积神经网络。它能够根据关联度对信息进行加权,让网络学习更重要的特征。2018 年 Hu 等^[21]提出 SE 模块,使

表 1 ResNet-18 结构表
Table 1 ResNet-18 structure table

Layers	Output Size	ResNet-18
Conv1	112×112	7×7 conv
MaxPool	56×56	3×3 max pool
Block(1)	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$
Block(2)	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$
Block(3)	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$
Block(4)	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$
GAP	1×1	Global Average Pooling
Fc		dense

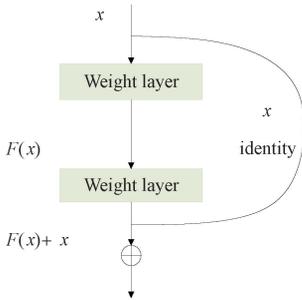


图 1 残差学习单元结构

Fig.1 Structure of residual learning unit

得网络能够自适应地学习每个特征通道的重要性。2019年提出的 SK 模块,通过自适应地选择和组合不同大小的卷积核,以捕获不同尺度下的特征信息。CBAM^[22]和 BAM^[23]模块,分别采用串行和并联的方式,学习通道与空间特征间的关联性和重要性。许晨扬等^[24]提出在多尺度特征图之间建立跨通道关联,实现一种多尺度通道注意力。Hou^[25]等提出 CA 注意力模块,分别沿着两个空间方向捕获长程依赖以及精确的位置信息进行互补以增强目标的表示。上述注意力模块虽然可以增强模型性能,但通常引入复杂的卷积或全连接操作,对于训练特征复杂的壁画图像,可能会增加计算和存储资源的消耗。随着 2020 年 ECA 注意力模块被提出,使用自适应的一维卷积局部跨信道交互的方式降低模块复杂度。王海勇等^[26]提出 ESCA 模块,克服了 ECA 模块无法获取图像空间信息的问题。但其空间注意力模块本质上还是基于池化操作,关注的是局部窗口内的特征,忽略了全局上下文信息,会导致图像的空间信息丢失。导致细节和局部特征的损失。本文方法对 ESCA 做进一步的改进,提出一种多尺度高效的空间通道注意模块 MESCA。内部使用 SK 注意力模块,提取图像的多尺度空间信息,捕捉全局长距离依赖关系。有助于更好地理解壁

画图像中目标位置、大小的相互关系,得到更具代表性的特征表达。

2 古代壁画分类模型设计

2.1 分类网络框架

基于改进 ResNet 深度学习的古代壁画分类方法(classification of ancient murals based on improved resnet deep learning, P-ResNet)是针对古代壁画图像分类任务而提出的,模型在 ResNet-18 的基础上进行优化,具体结构如图 2 所示。P-ResNet 模型的关键技术包括 3 部分:1)针对壁画纹理、边缘的浅层特征深度提取模块(SD-Model);2)为解决壁画背景噪声干扰提出一种多尺度高效的空间通道注意模块(MESCA);3)提出了一种蜂窝式聚合结构(HP-Structure)以区分不同场景下壁画人物特征差异。

首先对壁画图像进行预处理操作,将壁画图像裁剪到 224×224 尺寸,将每个壁画特征数据以三通道的 224×224 尺寸作为输入送入改进的特征提取网络进行壁画特征表示。第 1 阶段由 SD-Model 对输入图像进行卷积和池化操作,提取壁画图像中细微纹理特征,以及将目标对应的背景与弱纹理区域的特征提取出来。第 2 阶段是特征提取阶段,构建一种蜂窝式聚合结构,具体如图 2 所示。HP-Structure 由 4 个 Block 聚合而成,每个 Block 包含两层 Bottleneck Block Structure,可以综合壁画整体结构信息和细节信息;另外将 MESCA 注意力模块嵌入每个 Block 中可以获得全面而具有区分性的特征表示。第 3 阶段是由全局平均池化层、全连接层和 Softmax 层构成,特征图经过全局平均池化层得到每个特征图的全局权重,全连接层将全局平均池化层输出的特征向量作为输入,通过学习权重和偏置来将其映射到最终的类别或标签空间,通过 Softmax 函数计算出样本所属壁画标签的概率值,其中 \mathbf{z} 是一个向量, k 和 j 是其中的一个元素索引,公式如式(2)所示。

$$p(k|x) = \frac{\exp(\mathbf{z}_k)}{\sum_{j=1}^k \exp(\mathbf{z}_j)} \quad (2)$$

文中通过设计损失函数基于交叉熵来计算预测壁画标签与真实标签的近似程度,如式(3)所示,其中 $F(\mu, p)$ 表示真实标签向量 μ 与预测标签结果 p 的交叉熵, k 表示壁画类别数, $\hat{\mu}_k$ 表示第 k 类的真实概率。

$$F(\mu, p) = - \sum_{k=1}^k \log(p_k) \hat{\mu}_k \quad (3)$$

为了增强网络的泛化能力,避免标签的过高置信度,使用标签平滑(label smoothing, LS)交叉熵函数^[27]如式(4)所示。其中, $\hat{\gamma}_k$ 如式(5)所示,表示标签平滑处理后壁画所属类别概率; ϵ 为平滑系数,默认取值为 $\epsilon = 0.1$ 。

$$F^*(\mu, p) = - \sum_{k=1}^k \log(p_k) \hat{\gamma}_k \quad (4)$$

$$\hat{\gamma}_k = \hat{\mu}_k (1 - \epsilon) + \frac{\epsilon}{k} \quad (5)$$

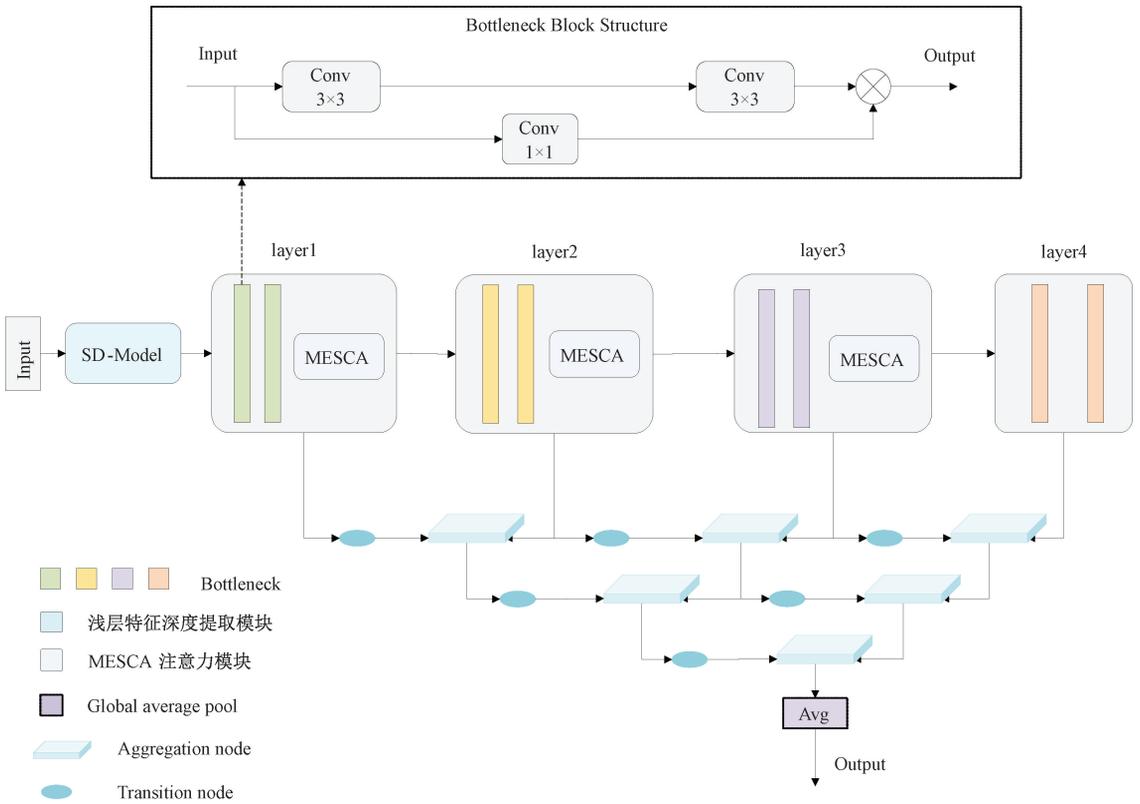


图 2 本文方法结构图

Fig. 2 Structure of the methodology in this paper

2.2 浅层特征深度提取模块

由于古代壁画不同人物之间纹理、轮廓、色彩极为相似,特征难以区分,在自动分类过程中容易混淆。因此本文设计了一种浅层特征深度提取模块,保留壁画图像中显著纹理特征,以及将目标对应的背景与弱纹理区域的特征提取出来,如图 3 所示,首先将模型输入层中较大的 7×7 卷积核分离为 3 个串联的 3×3 小卷积核堆积的主干,降低参数数量的同时,增强网络的非线性表达能力,更好地学习复杂的特征和模式。其次将 2×2 平均池化与最大池化进行 add 特征融合取代原最大池化操作。平均池化和最大池化分别捕捉了壁画图像中不同层次的信息。平均池化可

以平滑图像并保留整体趋势,有助于减少梯度消失或梯度爆炸问题。而最大池化可以捕捉到更明显的特征。通过将两者的结果相加,不仅可以增加壁画特征的多样性和丰富性,减少信息的损失,也可以使梯度传播更加平滑和稳定。最后将所有步长取值为 1,保留更多的空间信息,因为在每个方向上都只移动一个像素,不会因为步长过大而导致部分区域被忽略,使得输出特征图的大小与输入特征图相似,有助于保留更多的细节信息,生成更精细的特征表示,避免信息丢失和过度压缩。使得网络能够在更小的空间范围内捕捉壁画图像的局部模式和细微变化,这对于需要更多细节信息的壁画分类任务尤为重要。

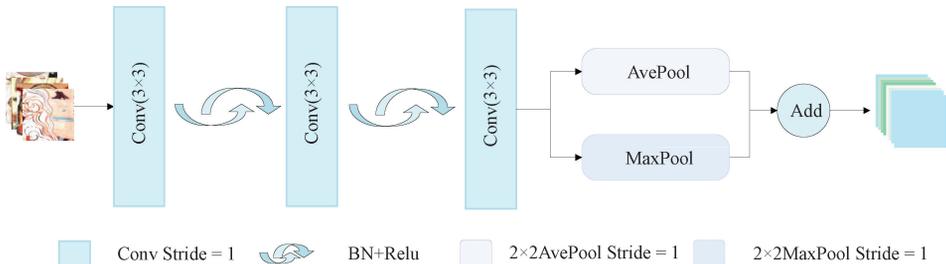


图 3 浅层特征深度提取模块结构

Fig. 3 Structure of shallow feature depth extraction module

2.3 多尺度高效的空间通道注意模块

本文 P-ResNet 模型针对壁画图像出现的褪色、剥落、物理损坏等而导致的背景噪声复杂,在特征提取阶段难以辨别有价值信息的问题,引入下述准则做了一个重要的设计——提出多尺度高效的空间通道注意力模块 (MESCA),聚焦关键信息,提升网络性能。MESCA 注意力模块嵌入特征提取网络前 3 个 Layer 层内,其结构如图 2 所示,输入的特征图首先经过两个 3×3 的卷积层,提取壁画抽象局部特征的同时实现空间尺寸的降维。再经过 BN、Relu 激活函数非线性变换,使网络可以学习到更加复杂的特征表示。而由于古代壁画图像具有较强的背景噪声以及壁画分类的对象往往包含于背景,无法很好地从背景中单独分割出来。因此在特征提取时,在面对背景噪声干扰的情况下保留壁画中人物的显著纹理特征,还应将目标区域对应的背景与弱纹理区域的特征进行区分。因此将卷积层提取的壁画特征再经过 MESCA 注意力模块

进行特征筛选,高效地获得全面而具有区分性的特征表示,抑制其他无用的背景噪声信息。最后经过一个残差学习单元融合原始输入信息形成最终的输出。

MESCA 注意力模块结构如图 4 所示,主要包括两部分:通道注意力模块和改进的空间注意力模块,其中通道注意力模块就是原始的 ECA 注意力模块,ECA 是使用大小为 k 的一维卷积代替全连接层在通道之间做卷积操作生成通道权值,能够根据感受野的不同自适应的调整其大小,其中卷积核 k 决定了交互的覆盖范围,并且 k 值与通道维度 C 成比例,通过通道维度 C 来自适应确定,克服了过去通道注意力模块复杂度过高的缺陷。通道维度 C 与卷积核大小 k 的关系如式(6)和(7)所示。其中,通道维度 C 一般设置为 2 的整数次幂,一般将 α 和 b 分别设为 2 和 1。

$$C = \sigma(k) = 2^{\alpha \times k - b} \tag{6}$$

$$k = \varpi(C) = \left\lfloor \frac{16C}{\alpha} + \frac{b}{\alpha} \right\rfloor_{\text{odd}} \tag{7}$$

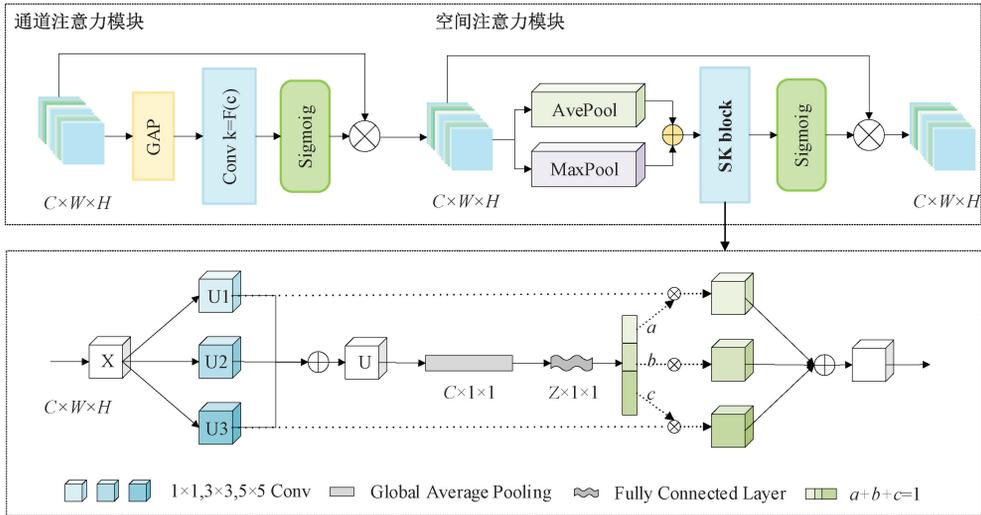


图 4 多尺度高效的空间通道注意力模块

Fig. 4 Multi-scale efficient spatial channel attention module

在 ECA 的基础上串联改进的空间注意力模块构建新的 MESCA 注意力模块。假设将高度、宽度、通道数分别为 W 、 H 、 C 的图像作为 MESCA 的输入,整体的过程如式(8)所示, f_{channel} 代表 ECA 通道注意力。改进的空间注意力的过程如式(9)所示,其中 x 表示上一层的输入, ψ_{max} 和 ψ_{min} 表示最大池化和平均池化操作, ζ_{SK} 代表 SK 注意力模块, τ 表示通道连接操作, σ 代表激活函数。原空间注意力模块本质上还是基于全局平均池化和全局最大池化的操作,之后通过全连接层,整合空间特征信息。在特征图压缩过程中会导致了大量信息的丢失,同时无法区分壁画图像中不同区域的不同特点,这可能导致模型在处理壁画图像中复杂的区域关系时表现不佳,因此本文采用 SK 模块替换原空间注意力全连接操作。SK block 如图 4 所示:首先将特征图 X 经过 1×1 、 3×3 、 5×5 卷积得到 U_1 、 U_2 、

U_3 三个特征图,然后相加得到了 U , U 中融合了多个感受野的信息,通过 global average pooling 操作,得到关于 channel 的一个 $C \times 1 \times 1$ 的一维向量,表示各个通道的信息的重要程度。接着用线性变换,将原来的 C 维映射成 Z 维的信息再还原为原来的 C 维,之后使用 Softmax 进行归一化,计算每一个卷积通道的重要性,将其进行信息融合,得到最终输出。利用 SK 模块通过不同尺度的卷积核对不同大小的目标捕获各通道及其相邻通道之间的交互信息得到最合适的特征表示,通过融合多尺度信息捕捉全局长距离依赖关系,有助于更好地理解壁画图像中目标位置、大小和相互关系,使空间注意力模块能够更加精准的筛选对壁画分类有积极作用的特征。

$$x_{\text{out}} = f_{\text{channel}}(x_{\text{in}}) \cap f_{\text{spatial}}(f_{\text{channel}}(x_{\text{in}})) \tag{8}$$

$$f_{\text{spatial}}(x) = \sigma(\tau(\zeta_{\text{SK}}(\psi_{\text{max}}(x_{\text{in}}), \psi_{\text{avg}}(x_{\text{in}})))) \tag{9}$$

深层网络的感受野比较大,特征图的分辨率低,空间几何特征细节缺乏,而壁画图像的纹理、细节信息主要表现在通道方向上,轮廓形状等整体信息则体现在空间通道上。因此在网络中嵌入 MESCA 注意力模块,在获取通道信息的同时,也关注空间区域,进一步提升古代壁画分类的性能。

2.4 蜂窝式聚合结构

古代壁画分类网络可以分为浅层网络和深层网络,浅层网络负责提取目标边缘等底层特征,主要包含一些普遍的壁画特征模式,如纹理、颜色和方向,这些模式通常不是全局的。而深层网络可以构建高级的语义信息,反映了壁画的特定信息,通常是全局信息。针对壁画在不同场景下

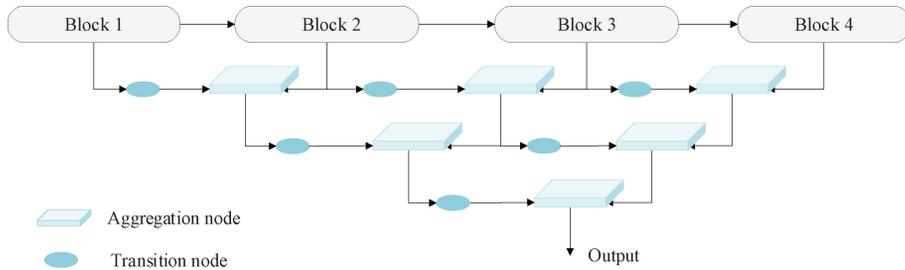


图 5 蜂窝聚合结构图

Fig. 5 Structure of honeycomb polymerisation

然后采用步长为 2 的 2×2 最大池化来对特征图进行下采样,压缩特征图的尺寸,获得了低维化的高级语义信息。Aggregation node 的级联操作受到 De-nseNet 密集连接的启发,获得前面层额外的输入,并将自己的特征映射传递到后续的所有层,生成融合特征,使浅层网络能够从所有前面层获得更多的输入,特别是深层网络的高级语义信息,从而在信息识别过程中获得更大的特征利用空间。通过 HP-Structure 结构,深层语义信息可以有效地流通到浅层网络中,极大缓解了传统网络中的“语义鸿沟”问题。语义鸿沟指的是浅层网络和深层网络之间信息传递不充分,导致浅层网络中无法有效学习到深层的全局信息。而 Aggregation Node 的级联操作则通过密集的特征流动,将上下文信息充分传递给浅层网络,综合壁画整体结构信息和细节信息,改善了这一问题。

3 实验结果与分析

3.1 实验环境与设计

实验设备:PC 环境为 Windows11 系统,CPU 采用 Intel Corei7-12700H, GPU 为 NVIDIA GEFORCE RTX3090,显存为 24 GB。使用基于 Python 语言的 Tensor-flow2.0 深度学习框架。壁画训练过程中采用 Adam 优化器反向传播更新网络参数,默认迭代轮数为 100 轮,实验以小批量的形式进行,batch size 设置为 12。设置学习率初始值为 0.000 1,采用学习率衰减策略,在第 50 轮将学习率衰减为原来的 1/10,其余均不发生改变。

人物特征差异较大,因此设计了一种蜂窝式聚合结构(HP-Structure),利用壁画高层次特征和低层次特征,通过这种结构,让深层网络更高级语义信息能够融合到稍浅层的网络,指导浅层网络进行信息识别。如图 5 所示,HP-Structure 由这 4 个 block 组成,block 之间通过非线性连接聚合了不同阶段的特征,而 transaction node 位于相邻 block 之间采用 1×1 卷积通过组合特征图中的通道来生成新的特征表示,通过这种方式,深层网络提取到的高级语义特征可以通过卷积重新编码成对浅层网络有用的低维表示。这种通道重组让浅层网络能够从深层特征中学到更多有助于局部信息识别的内容,提升其对细节的感知能力。

古代壁画图像数据集素材均选自《中国丝绸之路的墓室壁画》和《中国敦煌壁画全集》画册扫描后的作品,覆盖了南北朝、西魏、北周、隋、唐等多个盛世朝代的艺术图像,反映了古代社会的文化习俗和价值观念,具有独特的艺术风格和特色。在相关专家的指导下,根据壁画的主题和内容特点,同时考虑对分类任务的实际需求,分别选取了 8 种典型代表类别:世俗人、植物、菩萨、动物、建筑、祥云、佛弟子、佛,将其绘制成壁画数据集。并采用数据增强的方式如:水平和垂直翻转、亮度变换、加噪(高斯噪声和椒盐噪声)等进行数据扩充,并将其分为两组用于训练和测试,具体分配数量如表 2 所示,部分壁画数据集图像如图 6 所示。

表 2 数据集类别分配表

Table 2 Table of data set category assignments

类别	训练集	测试集	总数
世俗人	1 361	341	1 702
植物	1 195	299	1 494
菩萨	1 275	319	1 594
动物	1 343	336	1 679
建筑	996	249	1 245
祥云	1 108	278	1 386
佛弟子	1 072	269	1 341
佛	950	238	1 188

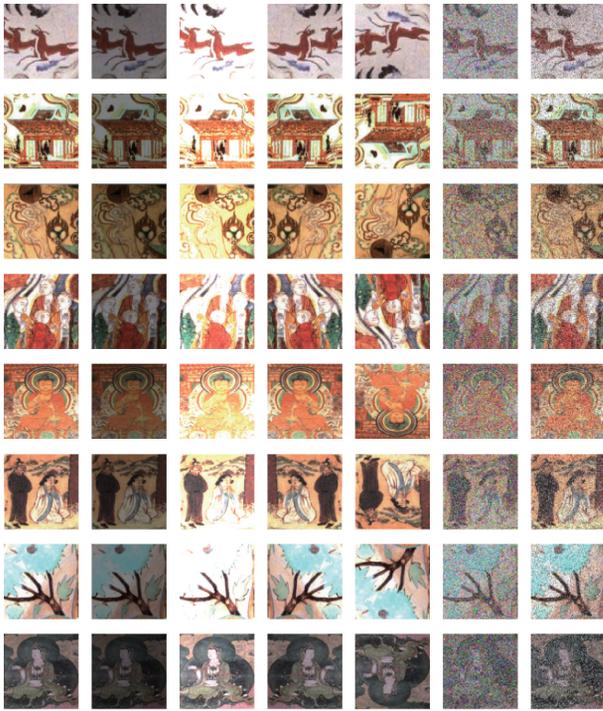


图 6 壁画数据集示例

Fig. 6 Example of mural dataset

3.2 评估指标

对于壁画数据集的单标签分类任务,本文使用准确率($A_{accuracy}$)、召回率(R_{recall})、精度($P_{precision}$)、F1 值($F_{F1-score}$) 4 个指标对分类效果进行评价,衡量分类模型的性能。具体计算公式如下:

$$A_{accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (10)$$

$$R_{recall} = \frac{TP}{TP + FN} \quad (11)$$

$$P_{precision} = \frac{TP}{TP + FP} \quad (12)$$

$$F_{F1-score} = \frac{2 \times P_{precision} \times R_{recall}}{P_{precision} + R_{recall}} \quad (13)$$

其中, TP 表示模型正确预测为正例的样本数, TN 表示模型正确预测为负例的样本数, FP 表示模型错误地将负例预测为正例的样本数, FN 表示模型错误地将正例预测为负例的样本数。这 4 个指标能够全面评估模型的分类性能, 准确率衡量了整体的正确率, 召回率衡量了正例的查全率, 精度衡量了正例的查准率, 而 F1 值则综合考虑了精度和召回率的平衡情况。

混淆矩阵提供了一种可视化方法, 展示了模型在每个类别上的预测结果, 并且可以用于计算各种分类指标, 可以帮助直观理解不同类别的错误预测情况。

3.3 与经典卷积神经网络对比分析

本次研究选取当下经典的多种卷积神经网络进行实验对比, 以此验证本文方法的优越性。选取经典的

AlexNet^[7]、GooleNet-V3^[10]、VggNet^[8]、DenseNet-121^[12]、EfficientNet-v2^[14]。使用上述网络模型对古代壁画数据集进行训练并记录数据, 各个模型在壁画数据集上准确率、精度、召回率、F1 值对比结果如表 3 所示。不同算法实验准确率变化曲线如图 7 所示, 其中横坐标代表迭代轮次, 纵坐标代表准确率值。

表 3 经典算法性能对比

Table 3 Performance comparison of classical algorithms

算法	准确率	精度	召回率	F1
AlexNet	87.57	88.20	87.67	87.79
GooleNet	76.95	77.71	76.58	76.84
VggNet	85.68	86.39	85.94	86.07
DenseNet	87.01	87.62	87.29	87.28
EfficientNet	83.27	84.29	83.77	83.90
本文	96.51	96.65	96.67	96.63

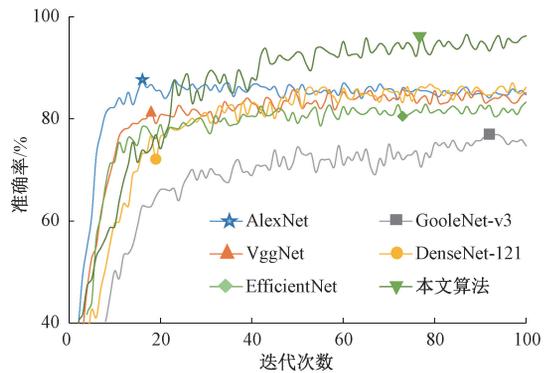


图 7 各经典算法准确率变化曲线

Fig. 7 Variation curve of accuracy of each classical algorithm

由表 3 可以看出, 本文方法达到了 96.63% 的准确率, 较 AlexNet 提高了 8.84%, 较 GooleNet-v3 提高了 19.76%, 较 VggNet 提高了 10.56%, 较 DenseNet-121 提高了 9.35%, 较 EfficientNet-v2 提高了 12.73%。其余 3 项指标: 精度、召回率、F1 值对比结果也显示本文算法较其他经典网络均有不同程度的提升。究其原因主要是浅层特征深度提取模块能够提取更多样和更丰富的信息, 生成更精细的特征表示; 通过 MESCA 注意力模块再捕捉特征的全局长距离依赖关系, 高效地获得全面而具有区分性的特征表示; 最后通过网络的蜂窝式聚合结构综合壁画整体的结构信息和细节信息, 充分表达本文壁画图像。结合图 7 来看, 本文算法较其他模型具有更高的初始准确率, 准确率也达到了最高。前期收敛速度相比略低, 主要原因是由于浅层特征深度提取模块提取的精细特征也包含了对分类无用的信息, 但模型中后期分类准确率远远高于其他模型, 主要是因为 MESCA 注意力模块突出了壁画重要特征, 而抑制其他无用信息, 可以使壁画图像得到更充分

的表达。

3.4 消融实验

1) 消融实验整体性能对比分析

为了验证不同模块的有效性,本节设计了多组对比实验。在 ResNet-18 网络嵌入浅层特征深度提取模块(SD-Model),简称 ResNet+SD,在 ResNet-18 网络嵌入多尺度高效的通道注意力模块(MESCA),简称 ResNet+MESCA。在 ResNet-18 网络嵌入蜂窝式聚合结构(HP-Structure),简称 ResNet+HP。同时加入 SD-Model 和 MESCA 简称为 ResNet+SD+MESCA。且同时加入 SD-Model 和 HP-Structure 简称为 ResNet+SD+HP。同时加入 MESCA 和 HP-Structure 简称 ResNet+MESCA+HP。各模型对比性能得分如表 4 所示。

表 4 不同模块下的消融实验对比结果

Table 4 Comparative results of ablation experiments under different modules

算法	准确率	精度	召回率	F1
ResNet-18	86.75	87.43	86.97	87.13
ResNet+SD	90.54	91.30	90.88	91.02
ResNet+MESCA	88.77	89.65	88.68	89.03
ResNet+HP	87.96	88.41	88.01	88.16
ResNet+SD+MESCA	95.91	96.28	95.95	96.09
ResNet+SD+HP	91.48	92.21	91.73	91.91
ResNet+MESCA+HP	89.46	90.31	89.45	89.80
本文	96.51	96.65	96.67	96.63

通过表 4 可以看出,ResNet+SD、ResNet+MESCA、ResNet+HP 三种模型的综合准确率分别达到 90.54%、88.77%、87.96%,较原模型 ResNet-18 分别提升了 3.97%、2.02%、1.21%,精度、召回率、F1 值也有不同程度的提升,说明在模型中单独加入 3 种结构都能够起到积极的效果。模型加入 SD-Model 能够生成更精细的特征表示,增强模型的特征能力。加入 MESCA 注意力能够通过融合多尺度信息捕捉全局长距离依赖关系,获得全面而具有区分性的特征表示,降低壁画背景噪声的干扰。加入 HP-Structure 能够同时捕获低级和高级特征,增强上下文信息的流通性。而 ResNet+SD+MESCA 和 ResNet+SD+HP 模型准确率分别达到 95.91%、91.48%,较 ResNet+SD 再次提升了 5.37%、0.94% 的性能,ResNet+MESCA+HP 模型准确率达到 89.46%,较 ResNet+HP 提升了 1.5%,以上说明任意两个模块通过相互作用都可以产生更好的效果。本文所提出的方法将 3 种结构进行融合,各性能分别达到 96.51%、96.65%、96.67%、96.63%,较其他模型各性能分别达到最高水平。究其原因主要是通过 SD-Model 提取了更精细的壁画特征表示,但其中也存在一些无用特征,但 MESCA 注意力能够将这些对壁画分类

无用的特征进行筛选区分并突出重要信息,最后通过 HP-Structure 综合壁画整体结构信息和细节信息,提高古代壁画模型的综合性能。

2) 浅层特征深度提取模块效果验证

为了验证浅层特征深度提取结构(SD-Model)的优越性,本节设计了此次对比实验。本文模型使用的 SD-Model 结构如图 3 所示。另一种移除平均池化,仅采用最大池化操作,将其命名为 ResNet-Ave。在此基础上将 3 个串联的 3×3 小卷积核堆积的主干还原为 7×7 卷积核,将其命名为 ResNet-None。实验结果如表 5 所示。

表 5 浅层特征深度提取结构实验对比

Table 5 Experimental comparison of shallow feature

算法	depth extraction structures			%
	准确率	精度	召回率	F1
本文	96.51	96.65	96.67	96.63
ResNet-Ave	94.96	95.23	95.09	95.11
ResNet-None	89.46	90.31	89.45	89.80

由表 5 可知,将 SD-Model 变换为其他两种形态的网络结构,最终实验结果均不及本文模型。本文算法较 ResNet-Ave 和 ResNet-None 准确率分别提升了 1.55%、7.05%,其余各性能也均有不同程度的提升。实验结果说明采用平均池化与最大池化进行特征融合能够取得更好的分类效果,原因在于平均池化可以平滑壁画图像并保留整体趋势,而最大池化可以捕捉到更明显的纹理特征。通过融合,可以增加特征的多样性以及丰富性。ResNet-Ave 模型较 ResNet-None 准确率提升 5.5%,说明步长为 1 的卷积主干可以避免壁画信息丢失和过度压缩,提高模型的性能和表现力。

混淆矩阵也可以作为分类模型评判指标,混淆矩阵的行、列标签分别代表壁画图像的预测与真实类别。行列交叉处的数值表示每个壁画类别被正确分类的数量。图 8 和图 9 分别为 ResNet-None 网络与本文模型的混淆矩阵。从图中可以看出对于 ResNet-None,采用 SD-Model 在壁画图像分类中误识率更低,且模型可以降低信息损失的可能性更准确地分辨壁画图像的细微差别。本文算法的混淆矩阵表现出更好的性能,更精准的识别效果。

3) 注意力机制性能对比分析

在本文模型中分别嵌入不同的注意力模块,其余不发生改变,以此验证 MESCA 注意力的优越性。在壁画数据集上进行训练,实验数据如表 6 所示。

由表 6 可以看出,嵌入 SE、ECA 与 ESCA 注意力对于壁画识别性能的提升显著,为壁画分类准确率贡献了 3.91%、4.04%、4.43% 的提升。而加入 CBAM、BAM、CA 注意力模块同样得到 3.87%、3.7%、3.27% 的提升。证明了网络在处理不同纹理、线条的壁画图像时,注意力模块



图 8 ResNet-None 混淆矩阵

Fig. 8 ResNet-None confusion matrix

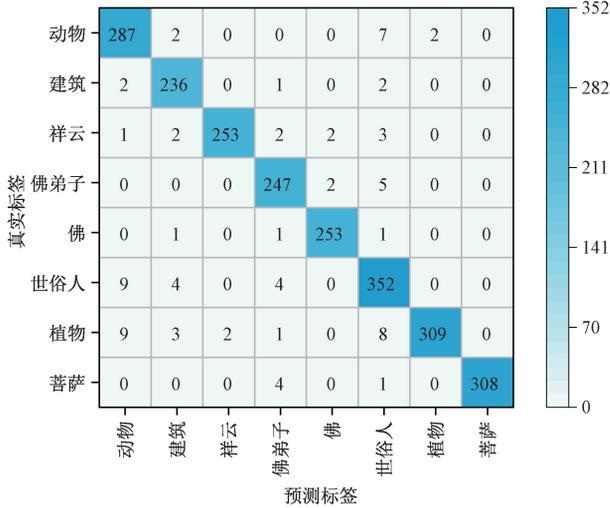


图 9 本文模型混淆矩阵

Fig. 9 Confusion matrix of the model in this paper

对模型赋予了聚焦关键特征的能力。本文引入的 MESCA 模块是在 ESCA 模块基础上进行了优化。由表 6 可以看出相比 ESCA 模块、ECA 模块,加入 MESCA 注意力模块的模型使得壁画分类准确率又进一步得到 0.99%、0.6%

表 7 与其他改进的神经网络准确率实验对比

Table 7 Experimental comparison of accuracy with other improved neural networks

算法	动物	建筑	祥云	佛弟子	佛	世俗人	植物	菩萨	AVG
文献[18]	91.27	92.53	90.49	85.82	92.18	86.72	91.26	92.97	90.32
文献[17]	89.26	77.59	82.50	81.49	91.40	78.86	82.83	82.74	83.23
文献[28]	90.60	89.21	92.77	81.10	90.62	79.40	91.56	95.20	88.65
文献[29]	92.28	89.62	93.91	80.70	90.23	86.44	89.15	92.01	89.29
文献[30]	80.87	93.77	93.15	84.25	90.62	85.09	90.36	94.88	88.95
本文	96.30	97.92	96.19	97.24	98.82	95.39	93.07	98.40	96.51

表 6 加入不同注意力模块分类对比结果

Table 6 Comparison results of adding different attention

算法	module classifications				%
	准确率	精度	召回率	F1	
None	91.48	92.21	91.73	91.91	
SE	95.39	95.86	95.31	95.55	
CBAM	95.35	95.87	95.40	95.58	
BAM	95.18	95.52	95.24	95.35	
CA	94.75	95.28	94.98	95.06	
ECA	95.52	96.04	95.53	95.71	
ESCA	95.91	96.14	95.99	96.03	
MESCA	96.51	96.65	96.67	96.63	

的提高,在多种注意力模块中对于壁画分类性能提升最为显著。对比表 6 第一行无任何注意力作用的网络模型,准确率、精度、召回率、F1 值分别获得 5.03%、4.44%、4.94%、4.72%的提升。

究其原因这是由于 MESCA 注意力模块采用不同大小的卷积核多尺度的捕捉全局长距离依赖关系,获得全面而具有区分性的特征表示。而 MESCA 模块相比于 CBAM 和 BAM 能够获得更好的性能,原因是 MESCA 使用原始的 ECA 模块采用局部跨通道交互方式获取通道权重,降低模型复杂度,并在空间维度多尺度的捕捉壁画特征间的内在关联,为空间像素信息赋予更精准的关注度,突出对壁画图像分类有决定性作用的空间信息。

3.5 与其他改进的神经网络实验对比分析

本文算法与代表性的壁画分类算法以及 ResNet 变体算法:文献[18]、文献[17]、文献[28]、文献[29]、文献[30]中的方法进行对比实验,对比结果如表 7 所示。由表 7 可知,Cao-F 与 Wang-N 古代壁画分类算法分类性能各方面较本文算法相对略低,这是因为本文算法的 SD-Model 模块能够提取壁画的细微特征,再通过 MESCA 注意力筛选、去除壁画通道与空间中的无用信息并加强重要特征,提升了模型对背景噪声的抗干扰性。其他改进的 ResNet 代表性分类算法对壁画的分类准确率最高达 89.29%,远低于本文算法分类性能。因为现有算法对于复杂多变的壁画特征不能够充分提取,并且壁画特征中包含许多噪声

信息,模型在分类过程中容易混淆,导致分类性能下降。本文算法嵌入 MESCA 注意力,获得全面而具有区分性的特征表示,之后通过 HP-Structure 综合壁画整体结构信息和细节信息,能够很好的解决这一问题。壁画图像分类各性能柱形对比如图 10 所示。由图 10 可知,本文模型在壁画分类任务上较其他代表性模型,在准确率、精度、召回率和 F1 值各方面都具有更好的性能。综合以上实验结果,这充分证明了本文模型在壁画分类任务中的先进性。

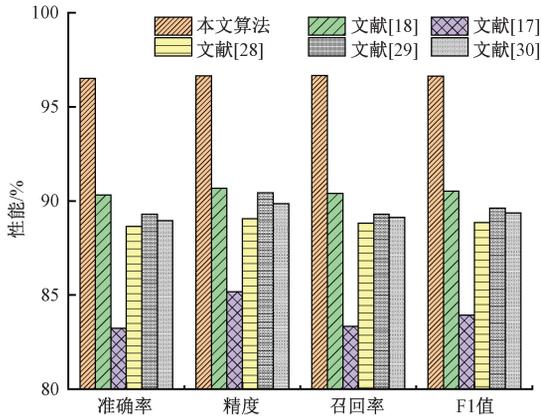


图 10 各算法性能对比图

Fig. 10 Comparison of the performance of the algorithms

4 结 论

本文设计了一个 P-ResNet 古代壁画分类模型。首先提出一种浅层特征深度提取结构,能够有效避免特征图信息丢失和过度压缩,并生成更精细的特征表示。其次在网络深层嵌入 MESCA 注意力模块,通过融合多尺度信息捕捉全局长距离依赖关系,获得全面而具有区分性的特征表示,降低背景噪声的干扰。最后通过蜂窝式聚合结构综合壁画整体结构信息和细节信息,缓解“语义鸿沟”问题。本文提出的方法有效解决了现有方法中因壁画不同人物之间特征相似,难以区分;背景噪声复杂;不同场景下壁画人物特征差异较大而导致分类性能不理想的难题。这对于文化遗产保护和艺术史方面研究具有显著价值。实验结果证明,本文算法分类准确度达到 96.51%,对比经典分类算法: AlexNet、GooleNet、VggNet、DenseNet、EfficientNet 分别提升了 8.84%、19.76%、10.56%、9.35%、12.73%,为后续研究提供了新的思路。

然而,本文网络模型仍存在一些问题,古代壁画中的图像可能具有多重含义或象征意义,这使得模型对图像进行准确的语义理解和分类变得更加困难。同时模型参数相对较大,计算量繁杂。因此,如何更好的解决壁画图像之间存在的“语义鸿沟”问题,设计一种轻量化的卷积网络模型是本文下一步的工作。

参考文献

[1] 陈娅敏,王可,王展,等. 基于多尺度超像素分割的壁画

多光谱图像颜料分类方法[J]. 激光与光电子学进展, 2024, 61(18): 478-488.

CHEN Y M, WANG K, WANG ZH, et al. A pigment classification method for multi-spectral images of murals based on multi-scale super pixel segmentation [J]. Advances in Laser and Optoelectronics, 2024, 61(18): 478-488.

[2] 郝亚博. 基于风格特征的中国古代人物壁画分类算法研究与实现[D]. 天津: 天津大学, 2017.

HAO Y B. Research and implementation of classification algorithm for ancient Chinese figure murals based on stylistic features[D]. Tianjin: Tianjin University, 2017.

[3] 钱文华,徐丹,徐瑾,等. 基于信息熵的风格绘画分类研究[J]. 图学学报, 2019, 40(6): 991-999.

QIAN W H, XU D, XU J, et al. Artistic paintings classification based on information entropy[J]. Journal of Graphology, 2019, 40(6): 991-999.

[4] ZENG Z M, SUN SH Q, LI T T, et al. Exploring the topic evolution of Dunhuang murals through image classification [J]. Journal of Information Science, 2024, 50(1): 35-52.

[5] ZHANG Y M, SUN D X, YIN Y P, et al. Fast identification of mural pigments at Mogao Grottoes using a LIBS-based spectral matching algorithm[J]. Plasma Science and Technology, 2022, 24(8): 084003.

[6] RAWAT W, WANG ZH H. Deep convolutional neural networks for image classification: A comprehensive review[J]. Neural Computation, 2017, 29(9): 2352-2449.

[7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Association for Computing Machinery, 2017, 60(6): 84-90.

[8] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. Computer Science, 2014, DOI: 10.48550/arXiv.1409.1556.

[9] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016: 2818-2826.

[10] HE K M, ZHANG X Y, REN SH Q, et al. Deep residual learning for image recognition[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016.

[11] HUANG G, LIU ZH, MAATEN L V D, et al. Densely connected convolutional networks[C]. IEEE Conference on Computer Vision and Pattern

- Recognition, 2017: 4700-4708.
- [12] TAN M X, LE Q. EfficientNet: Rethinking model scaling for convolutional neural networks [C]. International Conference on Machine Learning. PMLR, 2019: 6105-6114.
- [13] TAN M X, LE Q. EfficientNetV2: Smaller models and faster training [C]. International Conference on Machine Learning, PMLR, 2021: 10096-10106.
- [14] 李丹, 王慢慢, 刘俊德, 等. 基于轻量级卷积神经网络的带钢表面缺陷识别[J]. 仪器仪表学报, 2022, 43(3): 240-248.
LI D, WANG M M, LIU J D, et al. Identification of surface defects on strip steel based on lightweight convolutional neural network[J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 240-248.
- [15] 张晋婧, 刘双峰, 丰雷, 等. 融合注意力机制的人脸识别算法研究[J]. 国外电子测量技术, 2023, 42(2): 107-113.
ZHANG J J, LIU S H F, FENG L, et al. Research on face recognition algorithm incorporating attention mechanism [J]. Foreign Electronic Measurement Technology, 2023, 42(2): 107-113.
- [16] LI Q Q, ZOU Q, MA D, et al. Dating ancient paintings of Mogao Grottoes using deeply learnt visual codes[J]. Science China Information Sciences, 2018, 61: 1-14.
- [17] 王燕妮, 朱丹娜, 王慧琴, 等. 基于卷积神经网络的壁画颜料多光谱图像分类[J]. 激光与光电子学进展, 2019, 56(22): 48-56.
WANG Y N, ZHU D N, WANG H Q, et al. Multi-spectral image classification of mural pigments based on convolutional neural network [J]. Laser and Optoelectronics Progress, 2019, 56(22): 48-56.
- [18] CAO J F, JIA Y M, CHEN H M, et al. Ancient mural classification methods based on a multichannel separable network [J]. Heritage Science, 2021, 9: 1-17.
- [19] WANG Q L, WU B, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE, 2020: 11531-11539.
- [20] LI X, WANG W H, HU X L, et al. Selective kernel networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 510-519.
- [21] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [22] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module [C]. European Conference on Computer Vision, 2018: 3-19.
- [23] PARK J, WOO S, LEE J Y, et al. BAM: Bottleneck attention module [C]. The 29th British Machine Vision Conference, 2019: 1-14.
- [24] 许晨扬, 范非易, 柯冠舟, 等. 基于多尺度通道注意力机制的行为识别方法[J]. 电子测量技术, 2023, 46(21): 114-122.
XU C H Y, FAN F Y, KE G Z H, et al. Human activity recognition method based on multi-scale channel attention mechanism [J]. Electronic Measurement Technology, 2023, 46(21): 114-122.
- [25] HOU Q B, ZHOU D Q, FENG J S H. Coordinate attention for efficient mobile network design [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [26] 王海勇, 潘海涛, 刘贵楠. 融合注意力机制和课程式学习的人脸识别方法[J]. 计算机科学与探索, 2023, 17(8): 1893-1903.
WANG H Y, PAN H T, LIU G N. A face recognition method integrating attention mechanism and course-based learning [J]. Computer Science and Exploration, 2023, 17(8): 1893-1903.
- [27] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2016: 2818-2826.
- [28] WANG J M, CHEN Q Y, SHI C H Y. Research on spider recognition technology based on transfer learning and attention mechanism [J]. Symmetry, 2023, 15(9): 1727.
- [29] ZHANG X H, LI H L, SUN S H H, et al. Classification and identification of apple leaf diseases and insect pests based on improved ResNet-50 model [J]. Horticulturae, 2023, 9(9): 1046.
- [30] MA X H, LI Z H W, ZHANG L. An improved ResNet-50 for garbage image classification [J]. Tehnicki Vjesnik, 2022, 29(5): 1552-1559.

作者简介

曹建芳(通信作者), 教授, 博士, 主要研究方向为数字图像理解、大数据技术等。

E-mail: kcxj122@126.com

彭存赫, 硕士研究生, 主要研究方向为深度学习、视觉信息处理。

E-mail: pchbyj@163.com

陈志强, 讲师, 博士, 主要研究方向为智能信息处理。

E-mail: 896775672@qq.com

杨卓林, 硕士研究生, 主要研究方向为深度学习、图像修复、视觉信息处理。

E-mail: yang2021090925@163.com