

一种基于多任务学习的文化创意陶瓷产品视觉特征提取方法

曹冰一¹, 吴晓彬², 徐梦溪², 樊尹莘³, 史文雅¹

¹江苏建筑职业技术大学设计创意学院, 江苏 徐州

²南京工程学院计算机工程学院, 江苏 南京

³南昌师范学院文学与历史学院, 江西 南昌

收稿日期: 2026年9月8日; 录用日期: 2026年9月28日; 发布日期: 2026年10月12日

摘要

文化创意陶瓷产品视觉特征提取, 即从图像中自动输出釉色、纹饰、器型及工艺风格等特征的类别标签, 是图像素材库自动标注以及基于特征的图像生成与检索等应用的基础性工作。传统方法依赖人工对这四类视觉特征逐一判断, 效率低且主观性强。现有深度学习方案采用单任务学习范式, 对每类特征分别训练独立的分类模型, 推理效率较低。本文提出一种基于多任务学习的文创陶瓷产品视觉特征提取方法 (MTL-CVFE), 将釉色、纹饰、器型、工艺风格四项任务整合于同一模型, 以单次前向传播同步输出类别标签。该方法以ResNeXt50-SSD提取基础特征, 引入扩张卷积支路增强纹饰感受野, 并设计四任务并行输出头, 实现四类特征的端到端同步识别。同时构建了相应标注数据集用于实验验证。结果表明, 本方法在测试集上综合平均精度均值达到85%以上, 在保持各任务识别精度的前提下, 推理效率较现有单任务深度学习方案有明显提升。本文方法输出的结构化特征向量可直接服务于设计素材库的自动标注与生成式设计模型的条件控制等下游应用场景。

关键词

文化创意陶瓷产品, 视觉特征提取, 多任务学习, 工业设计

A Multi-Task Learning Based Method for Visual Feature Extraction of Cultural and Creative Ceramic Products

Bingyi Cao¹, Xiaobin Wu², Mengxi Xu², Yinxin Fan³, Wenya Shi¹

¹School of Design Innovation, Jiangsu University of Architectural Technology, Xuzhou Jiangsu

²School of Computer Engineering, Nanjing Institute of Technology, Nanjing Jiangsu

³School of Literature and History, Nanchang Normal University, Nanchang Jiangxi

文章引用: 曹冰一, 吴晓彬, 徐梦溪, 樊尹莘, 史文雅. 一种基于多任务学习的文化创意陶瓷产品视觉特征提取方法[J]. 图像与信号处理, 2026, 15(4): 459-469. DOI: 10.12677/jisp.2026.154041

Abstract

Visual feature extraction of cultural and creative ceramic products, which automatically outputs category labels of features including glaze color, decorative pattern, form, and craft style from images, is a foundational task for applications such as automatic annotation of image material libraries and feature-based image generation and retrieval. Traditional methods rely on manual judgment of these four types of visual features, which is inefficient and highly subjective. Existing deep learning schemes adopt a single-task learning paradigm, training independent classification models for each feature type, resulting in low inference efficiency. This paper proposes a multi-task learning based method for visual feature extraction of cultural and creative ceramic products (MTL-CVFE), which integrates the four tasks of glaze color, decorative pattern, form, and craft style into a single model and outputs category labels synchronously through a single forward propagation. The method employs ResNeXt50-SSD for backbone feature extraction, introduces dilated convolution branches to enhance the receptive field for decorative patterns, and designs a four-task parallel output head to achieve end-to-end synchronous recognition of the four feature types. A corresponding annotated dataset was constructed for experimental validation. Results show that the proposed method achieves a comprehensive mean average precision (mAP) of over 85% on the test set, and under the premise of maintaining recognition accuracy for each task, the inference efficiency is significantly improved compared with existing single-task deep learning schemes. The structured feature vectors output by the proposed method can be directly applied to downstream scenarios such as automatic annotation of design material libraries and conditional control of generative design models.

Keywords

Cultural and Creative Ceramic Products, Visual Feature Extraction, Multi-Task Learning, Industrial Design

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

文化创意陶瓷产品的视觉特征包括釉色、纹饰、器型与工艺风格，是其艺术价值、文化属性与设计基因的核心载体[1]。釉色指陶瓷表面的呈色与质感，可分为青釉、白釉、红釉等类型；纹饰指绘制或模印于器物表面的装饰图案，如缠枝莲纹、云龙纹；器型指陶瓷制品的立体形态类别，如瓶、罐、碗等；工艺风格则指陶瓷制作过程中形成的综合性技术特征与审美范式，如青花、粉彩等。在文创陶瓷产品设计实践中，设计师需要按特征检索素材库、根据特征组合生成设计方案、或对大量样本进行特征分类以建立设计资源数据库，上述应用场景均需从陶瓷产品图像中提取结构化的视觉特征参数。这一视觉特征提取能力具体服务于两类下游任务：其一是设计素材库的批量自动标注，将无结构的图像集合转化为可按特征查询的结构化数据库；其二是生成式设计模型的条件输入，量化特征向量可作为扩散模型等生成方法的条件控制信号[2] [3]。

文化创意陶瓷产品图像特征提取的研究经历了从传统图像处理到深度学习的演进。早期工作基于手

工设计的特征描述子,使用颜色直方图、尺度不变特征变换和局部二值模式等方法对陶瓷釉色与纹理进行量化描述,再结合支持向量机等分类器进行识别[4]。这类方法在受控条件下取得了一定效果,但对图像质量敏感,泛化能力有限。

深度学习的引入显著提升了特征提取性能[5][6]。卷积神经网络(CNN)以其端到端的学习能力,自动从数据中提取层次化的特征表示,在多个陶瓷产品图像特征提取任务上超越了传统方法。例如,残差网络(ResNet)通过跨层连接缓解了深层网络的优化困难,能够稳定地提取陶瓷图像的抽象语义特征;轻量化网络(MobileNet)则适用于资源受限的部署场景。研究者已将各类 CNN 网络应用于釉色分类、纹饰识别与器型检测[7][8],验证了深度学习从陶瓷产品图像中提取有效视觉表征的可行性。

然而,上述研究多采用单任务学习范式,即一个模型对应一项特征提取任务[9]。当需要同时获取四类特征时,单任务模型部署方案在架构层面存在两个局限:(1) 需部署四个独立模型,每张图像经四次前向传播,推理延迟为各模型之和;(2) 各模型独立训练,未利用工艺风格、釉色与纹饰之间天然的统计相关性。此外,在工程实践中,人工提取方式仍广泛使用,但人工提取存在速度慢、主观性强、难以规模化实施等局限。

针对上述单任务模型与人工提取的局限,本文提出一种基于多任务学习的文化创意陶瓷产品视觉特征提取方法(简称: MTL-CVFE),将釉色分类、纹饰分类、器型分类与工艺风格分类四项特征提取任务整合于同一模型。核心思路为:以 ResNeXt50-SSD 深度网络作为基础特征提取网络[10][11],构建四任务并行输出头,使单个模型在一次前向传播中同步输出四类特征的识别结果及量化特征向量[12]。四项特征提取任务共享基础特征提取网络,在优化过程中通过共享特征提取层实现信息交互,起到协同学习的效果。

本文的主要贡献:(1) 提出一种基于多任务学习的文化创意陶瓷产品视觉特征提取方法(MTL-CVFE),将四项特征提取任务整合于单一模型,推理效率较单任务方案明显提升;(2) 针对陶瓷纹饰的大幅面结构特性,在 ResNeXt50-SSD 深度网络中引入扩张卷积支路,通过扩大感受野增强模型对大尺度纹饰整体结构的感知能力[13]。本文方法处于视觉特征提取的上游环节,其输出的结构化特征向量既可直接用于设计素材库的自动标注,也可作为生成式设计模型的条件控制信号。

本文后续安排如下:第2节阐述系统整体架构;第3节详述关键问题与对应技术方案;第4节报告实验设置与结果分析;第5节总结全文。

2. MTL-CVFE 的整体流程设计

本文所提方法(MTL-CVFE)的整体处理流程由三个环节串联构成:输入预处理、特征提取与结构化输出。输入图像经预处理后送入特征提取网络,以单次前向传播同步完成四类特征的识别,最后由输出层将识别结果汇总为结构化的特征向量。整体流程如图1所示。

2.1. 输入层

输入层接收待识别的陶瓷产品图像,完成图像的预处理操作。输入图像为 RGB 彩色图像,尺寸统一缩放至 512×512 像素。预处理包括两个步骤:(1) 色彩空间标准化,将所有图像转换至 sRGB 色彩空间;(2) 图像归一化,将像素值从 $[0, 255]$ 映射至 $[0, 1]$ 区间。预处理后的图像张量 $X \in \mathbb{R}^{512 \times 512 \times 3}$ 送入特征提取层。

2.2. 特征提取层

特征提取层为系统的核心模块,以 ResNeXt50-SSD 为基础特征提取网络[10][11],并在其之上附加两个辅助结构:扩张卷积支路与四任务并行输出头。该层以单次前向传播完成四类特征的同步提取,由三个子模块串联组成。

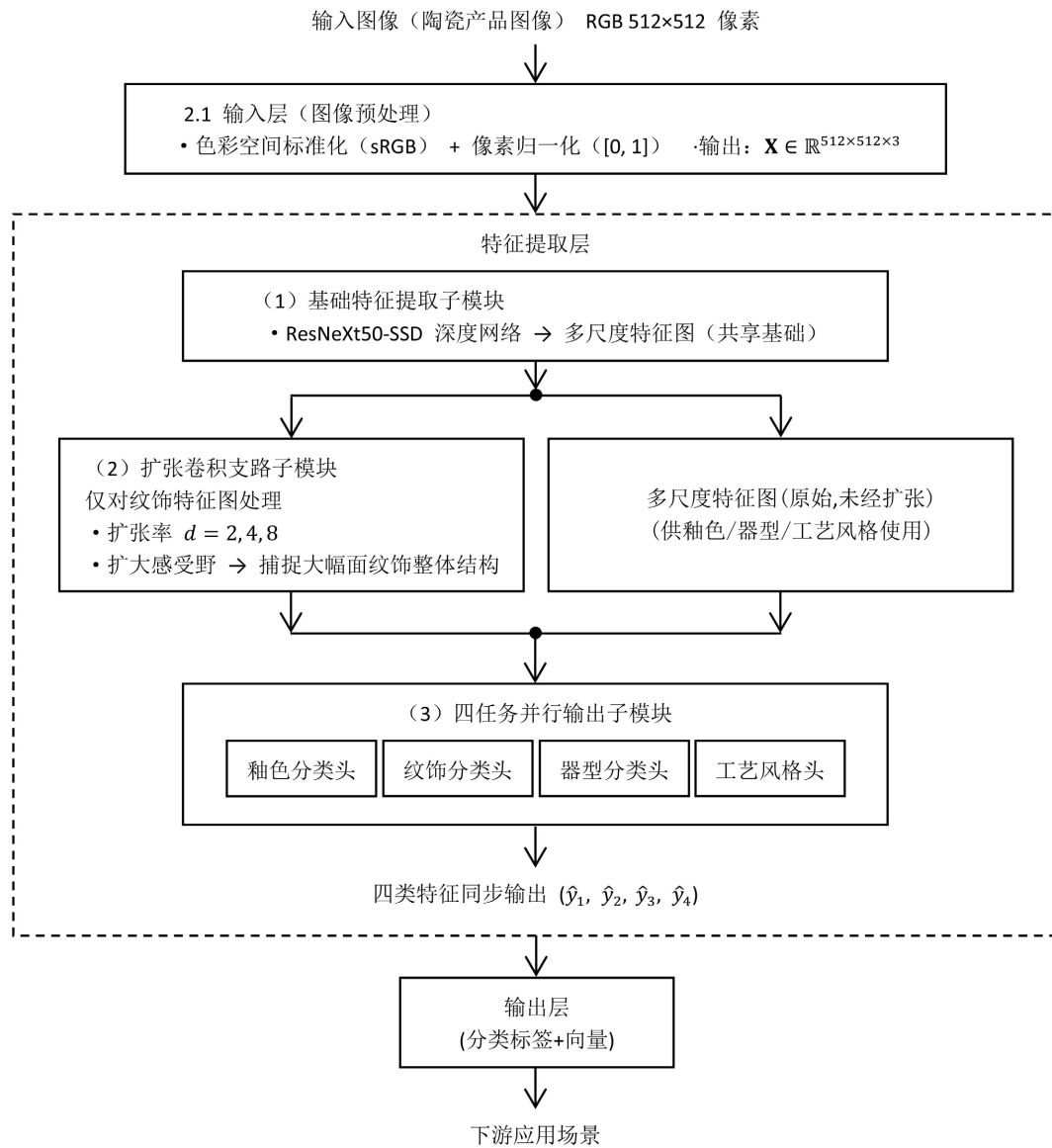


Figure 1. MTL-CVFE overall process

图 1. MTL-CVFE 整体流程

(1) 基础特征提取子模块。以 ResNeXt50-SSD 为主干网络，对输入图像进行多层卷积与池化操作，输出多尺度特征图。该子模块负责从输入图像中提取通用的视觉表征，为后续四项特征提取任务提供共享的底层特征基础。

(2) 扩张卷积支路子模块。该子模块仅作用于纹饰特征提取任务，在基础特征提取网络最高层特征图上，以扩张率 $d = 2, 4, 8$ 分别进行三层扩张卷积操作，用于扩大感受野以捕捉缠枝莲、云龙纹等大幅面纹饰的整体结构信息[13]。其余三项任务(釉色、器型、工艺风格)不使用该支路。各扩张率卷积的输出沿通道维度拼接后，经 1×1 卷积降维，与原始 SSD 特征图拼接，共同输入纹饰分类头。

(3) 四任务并行输出子模块。设置四个独立的分类头，分别对应釉色、纹饰、器型与工艺风格四项特征提取任务。其中，釉色、器型、工艺风格三个分类头接收基础特征提取网络输出的原始多尺度特征图；纹饰分类头接收经扩张卷积支路处理后的特征图。各分类头在高层进行任务特化的分类决策[9][12]。

基础特征提取子模块与扩张卷积支路子模块共同构成共享的基础特征提取网络部分，四任务并行输出子模块为任务特化的决策部分。共享与特化相结合的设计，使模型在保持高效率的同时兼顾各任务的差异化需求。

2.3. 输出层

输出层将四类特征的识别结果汇总为结构化的输出格式。输出包含两部分内容。

(1) 分类标签。四类特征的预测类别标签 $\hat{y}_1, \hat{y}_2, \hat{y}_3, \hat{y}_4$ ，分别对应釉色、纹饰、器型与工艺风格的分类结果。

(2) 量化特征向量。取基础特征提取网络输出的多尺度特征图，经全局平均池化后的向量 $\mathbf{v} \in \mathbb{R}^{2048}$ ，作为输入图像的量化特征向量。该向量是对图像视觉内容的紧凑化数值表征。

量化特征向量的具体形式为 2048 维浮点数向量。以一张青花缠枝莲纹梅瓶的图像为例，该向量长度为 2048，元素均为实数，取值范围大致在 -5 至 5 之间，是图像经深度网络层层抽象后的紧凑化表征。该向量与分类标签不同，分类标签是离散的类别名称(如“青花”“缠枝莲”“梅瓶”)，而量化特征向量是连续空间中的数值坐标，保留了图像视觉内容的丰富信息，适合作为数据库索引或生成模型的控制信号 [2] [3]。

输出层将分类标签与量化特征向量组合为结构化的特征描述向量，供下游场景直接调用。

3. 关键问题与技术方案

3.1. 关键问题分解

根据第 2 节描述的系统架构，特征提取层的设计面临以下三个关键技术问题。

问题一(Q1)：多任务共享基础特征提取网络的优化平衡问题。釉色、纹饰、器型、工艺风格四项特征提取任务共用同一基础特征提取网络。各任务在优化方向上可能存在差异，基础特征提取网络的优化可能偏向某一任务而损害另一任务。需要在特征共享与任务特化之间寻求平衡，使各任务的特征提取精度不因共享而显著下降。多任务学习的优势在于各任务可通过共享特征提取层实现协同学习 [9] [12]，但若权重分配不当，协同效应可能被任务间的竞争所抵消。

问题二(Q2)：纹饰大幅面结构的感受野不足问题。陶瓷纹饰中缠枝莲纹、云龙纹等类别具有大幅面结构特征，其形态信息分布在图像的大部分空间区域。标准卷积核的感受野有限，难以获取此类纹饰的整体结构信息，导致特征提取精度受限 [7] [8]。

问题三(Q3)：多任务提取的精度保障问题。多任务学习的理想效果是在提升推理效率的同时保持各任务的特征提取精度。需要设计合理的损失函数与训练策略，使多任务模型中每项任务的精度不低于对应的单任务专用模型。

3.2. 基础特征提取网络

本文选用 ResNeXt50-SSD 深度网络作为基础特征提取网络 [10] [11]。ResNeXt50-SSD 是一种专为图像识别与目标检测设计的复合神经网络架构，它将特征提取与多尺度目标检测集成于统一框架内。该网络由 ResNeXt50 与 SSD (Single Shot MultiBox Detector) 两个组件组合而成：ResNeXt50 通过分组卷积在保持参数量的同时提升了特征表达能力 [10]；SSD 是一种单阶段目标检测框架，其多尺度特征图设计可同时捕捉图像中不同尺寸的目标信息 [11]。该组合架构在陶瓷表面特征检测任务中已有应用验证 [8]，对陶瓷产品图像的纹理与色彩具有较好的表征能力。基础特征提取网络的输出特征图将分别输入扩张卷积支路(用于扩大纹饰感受野)和四个并行分类头(用于各任务分类决策)。

3.3. 问题一的技术方案：四任务并行输出头

针对问题 Q1, 本文设计四任务并行输出头结构。该结构在基础特征提取网络之上设置四个独立的分类头 h_1, h_2, h_3, h_4 , 分别对应釉色、纹饰、器型与工艺风格四项特征提取任务[9] [12]。

设基础特征提取网络输出的多尺度特征图为 $\{F_1, F_2, \dots, F_L\}$, 其中 LL 为 SSD 检测层的层数。对各尺度特征图分别进行全局平均池化后沿通道维度拼接, 得到综合特征向量 $F_{\text{concat}} \in \mathbb{R}^{C_{\text{total}}}$ 。四个分类头的输出为:

$$\hat{y}_k = h_k(F_{\text{concat}}) = \text{softmax}(W_k F_{\text{concat}} + b_k), k = 1, 2, 3, 4 \quad (1)$$

其中 W_k 和 b_k 为第 k 个分类头的权重矩阵与偏置项, 输出维度分别对应各类别的类别数。

四项特征提取任务共享基础特征提取网络的优势在于: (1) 单次前向传播即可完成全部四项特征提取任务, 推理效率显著高于四个独立模型的串联部署; (2) 共享底层特征使各任务可间接利用其他任务的学习信号, 起到隐式正则化的作用, 有利于提升泛化能力[9] [12]。各任务通过共享基础特征提取网络实现协同学习, 利用特征之间的统计相关性辅助彼此的分类决策。

3.4. 问题二的技术方案：纹饰专用的扩张卷积支路

针对问题 Q2, 本文在 SSD 检测层中增加一条扩张卷积支路, 专门用于扩大纹饰特征提取的感受野[13]。

该支路对基础特征提取网络最高层的特征图 $F_{\text{top}} \in \mathbb{R}^{H_{\text{top}} \times W_{\text{top}} \times C_{\text{top}}}$ 分别以扩张率 $d = 2, 4, 8$ 进行三层扩张卷积操作。扩张率为 d 的卷积核实际感受野大小为:

$$\text{RF}_d = k + (k-1)(d-1) \quad (2)$$

式中, k 为卷积核尺寸(本文取 $k=3$)。当 $d=1, 2, 4, 8$ 时, 等效感受野分别为 3×3 、 5×5 、 9×9 、 17×17 。

各扩张率卷积的输出沿通道维度拼接, 经 1×1 卷积降维后与原始 SSD 特征图拼接, 共同输入纹饰分类头。该改进仅作用于纹饰特征提取支路, 不影响其他三项任务的特征输入。

3.5. 问题三的技术方案：损失函数与训练策略

针对问题 Q3, 本文在损失函数设计和训练策略上进行专门处理, 以保障各任务的特征提取精度[9] [12]。

总损失为各任务交叉熵损失 $\mathcal{L}_k$ 的加权和:

$$\mathcal{L}_{\text{total}} = \sum_{k=1}^4 \lambda_k \mathcal{L}_k \quad (3)$$

权重 λ_k 依据各任务训练集的样本量进行自适应设定。样本量少的任务赋予较高权重, 样本量多的任务赋予较低权重:

$$\lambda_k \propto \frac{1}{\log N_k} \quad (4)$$

式中, N_k 为第 k 项任务的训练样本数。此策略可缓解各任务间样本量差异所导致的优化偏向。

训练采用迁移学习策略。基础特征提取网络使用在 ImageNet 上预训练的权重进行初始化, 分类头为随机初始化[5] [6]。优化器选用 AdamW, 初始学习率设为 1×10^{-4} , 采用余弦退火调度至 1×10^{-6} 。训练在 NVIDIA RTX 4090 GPU 上完成, batch size 设为 32。

4. 实验设计与结果分析

本节通过对比实验, 验证本文方法(MTL-CVFE)在文创陶瓷产品视觉特征提取任务上的精度提升效

果与推理效率优势。

实验流程包含训练与测试两个阶段。训练阶段，标注数据集按比例划分为训练集、验证集与测试集。训练集用于更新网络参数，使模型学习从图像到四类特征标签的映射关系；验证集用于监控训练过程中的过拟合程度，并选取性能最优的模型参数。测试阶段，将测试集图像输入训练完成的模型，输出四个分类头的预测标签，与人工标注的真实标签比对，计算精度、召回率与 mAP 等评估指标。测试集不参与训练过程，其评估结果反映模型对未参与训练的陶瓷产品图像的特征提取泛化能力[14]-[16]。

4.1. 实验数据

本研究构建了文创陶瓷产品视觉特征标注数据集，用于模型的训练与性能评估。该数据集包含约 1000 张陶瓷产品图像，覆盖釉色(10 类)、纹饰(12 类)、器型(7 类)、工艺风格(5 类)四类视觉特征。在陶瓷产品图像识别领域，公开可用的标准数据集相对稀缺。现有数据集多针对单一特征(如仅釉色分类或仅器型识别)构建，或来源于特定博物馆馆藏且获取受限，难以满足多任务学习模型对四类特征同步标注的数据需求。为此，本研究自建了面向四类特征同步标注的数据集，以支撑本文方法的实验验证。

图像来源于三类渠道：公开文化遗产数据库(约 40%)、拍卖与收藏机构图录(约 35%)、田野采集与教学标本(约 25%)。所有图像统一缩放至 512 × 512 像素，色彩空间统一转换为 sRGB [8]。

标注工作由 3 位标注人员完成，均具备陶瓷设计或工艺相关专业背景。标注流程包括独立标注、一致性检验与专家仲裁三个阶段[1]。标注人员独立标注，互不干扰。标注不一致的样本由第三名仲裁者最终裁决。Fleiss’ Kappa 系数 $\kappa > 0.80$ ，超过预设合格阈值(>0.70)，标注一致性达到可接受水平。四类特征的类别体系见表 1。

数据集按 7:1.5:1.5 的比例划分为训练集(约 700 张)、验证集(约 150 张)与测试集(约 150 张)。本数据集为面向研究验证的专用数据集，仅用于验证本文所提方法的有效性。

本数据集以四类特征的类别标签为标注目标。模型训练与测试均基于此标注体系：输入为陶瓷产品图像，输出为四类特征的预测类别标签，同时输出一个 2048 维的量化特征向量，该向量由基础特征提取网络输出的多尺度特征图经全局平均池化得到，元素均为浮点数。分类标签用于评估识别精度，量化特征向量用于下游任务对接。

Table 1. Classification system of visual features of cultural and creative ceramic products
表 1. 文创陶瓷产品视觉特征类别体系

特征维度	类别数	类别列表
釉色	10	青釉、白釉、红釉、蓝釉、黄釉、绿釉、黑釉、窑变釉、结晶釉、无釉/素胎
纹饰	12	缠枝纹、云龙纹、花鸟纹、山水纹、人物故事纹、几何纹、冰裂纹/开片、窑变纹理、诗文书法、莲瓣纹/蕉叶纹、海水纹、素面/无纹饰
器型	7	瓶类、罐类、碗类、盏/杯类、盘/碟类、壶类、文房杂项
工艺风格	5	青花、粉彩/五彩、颜色釉(单色釉)、釉下彩绘、雕塑/堆塑

4.2. 实验环境与评估指标

实验硬件配置为：NVIDIA RTX 4090 GPU (24 GB 显存)；Intel Core i9-13900K CPU；64GB DDR5 内存。软件框架为：Ubuntu 22.04 LTS 操作系统；PyTorch 2.0.1 深度学习框架；CUDA 11.8 + cuDNN 8.9.0。

实验在一台配置 NVIDIA RTX 4090 GPU (24 GB 显存)的计算服务器上完成，CPU 为 Intel Core i9-13900 K，内存为 64 GB DDR5。软件环境为 Ubuntu 22.04 LTS 操作系统、PyTorch 2.0.1 深度学习框架及

CUDA 11.8+cuDNN 8.9.0 加速库。训练过程中,所有图像输入尺寸统一为 512×512 , batch size 设为 32, 最大训练轮数为 60, 初始学习率为 1×10^{-4} , 采用余弦退火调度至 1×10^{-6} 。

评估指标采用精度(Precision)、召回率(Recall)和平均精度均值(mean Average Precision, mAP)三项标准指标[7] [8]。

精度定义为:

$$P = \frac{TP}{TP + FP} \quad (5)$$

召回率定义为:

$$R = \frac{TP}{TP + FN} \quad (6)$$

mAP 为各类别平均精度的均值, 综合反映模型在多类别分类任务上的整体性能。

4.3. 基线方法

为验证本文方法的有效性, 设置以下基线方法[8]。

(1) 单任务基准。针对四类特征分别训练四个独立的 ResNeXt50-SSD 分类模型, 各模型基础特征提取网络相同, 输出头为单任务结构[7] [8]。用于验证多任务学习相对于单任务部署的精度差异。

(2) 标准多任务。采用本文的四任务并行输出结构, 但不包含扩张卷积支路, 基础特征提取网络为标准 ResNeXt50-SSD。用于验证扩张卷积支路对纹饰特征提取精度的贡献[9] [12]。

(3) 本文方法(MTL-CVFE)。包含四任务并行输出头与扩张卷积支路的完整模型。

4.4. 特征提取实验结果

表 2 对比了各方法在测试集上的特征提取性能。测试集中每张图像均有四类特征的人工标注标签作为真实值。各方法以图像为输入, 经前向传播后输出四个分类头的预测标签, 将预测结果与真实标签比对, 计算出表中所列的各项 mAP 指标。推理时间记录从图像输入到四类特征标签全部输出所需的端到端耗时, 用于衡量各方法的实际运行效率。

结果分析如下[7] [8] [13]。

(1) 多任务学习与单任务基准的对比。标准多任务方法在釉色、器型、工艺风格三项上的 mAP 与单任务基准的差距均在 1 个百分点以内, 说明多任务共享基础特征提取网络未对多数任务造成显著的性能损失[9] [12]。纹饰任务的降幅相对较大(2.6 个百分点), 原因是纹饰特征提取对感受野的依赖高于其他任务, 这是本文引入扩张卷积支路的直接动机[13]。

(2) 精度差异的讨论。从表 2 可以看出, 本文方法在釉色、器型、工艺风格三项任务上的 mAP 与单任务基准的差距均在 1 个百分点以内, 纹饰任务上本文方法(85.1%)与单任务基准(85.2%)基本持平。这种差异的本质在于: 四项特征提取任务共享同一基础特征提取网络, 网络参数的优化方向是四项任务损失的加权综合, 而非针对某一单项任务的最优解, 因此在单项任务的极端优化性能上略低于专用模型, 这是多任务学习范式的固有特性[9] [12]。对于设计素材库自动标注等应用场景, 85%以上的精度配合 3.8 倍的效率提升, 微小的精度差异完全可以接受; 但对于文物真伪鉴定等要求极高精度的场景, 则需结合人工复核或采用专用模型。

(3) 扩张卷积支路的效果。本文方法(MTL-CVFE)在纹饰特征提取上的 mAP (85.1%)较标准多任务(82.6%)提升 2.5 个百分点, 验证了扩张卷积支路通过扩大感受野改善大尺度纹饰特征提取的有效性[13]。釉色、器型、工艺风格三任务的精度与标准多任务基本持平, 说明扩张卷积支路未对其他任务产生干扰。

(4) 推理效率。本文方法的推理时间为 33 ms/张, 单任务基准为 125 ms/张(四个模型合计)。在综合 mAP 相近的前提下, 推理效率提升约 3.8 倍。

Table 2. Comparison of feature extraction performance of various methods on the test set

表 2. 各方法在测试集上的特征提取性能对比

方法	釉色 mAP (%)	纹饰 mAP (%)	器型 mAP (%)	工艺风格 mAP (%)	综合 mAP (%)	推理时间(ms/张)
单任务基准	86.5	85.2	90.8	87.3	—	125
标准多任务 (无扩张卷积)	85.9	82.6	90.2	86.7	86.4	31
本文方法 (完整模型)	85.8	85.1	90.0	86.5	86.9	33

各特征维度上的详细性能如表 3 所示。表 3 进一步展示本文完整模型在四个子任务上的详细性能, 将表 2 中本文方法的测试结果按特征维度拆解为精度、召回率与 mAP 三项指标。通过对比四个子任务的 mAP 可判断模型在各类特征上的识别能力差异; 通过对比同一任务中精度与召回率的数值关系可判断模型是否存在预测偏向(精度高召回率低表示保守预测, 召回率高精度低表示积极预测)。

Table 3. Detailed feature extraction performance on each feature dimension (method proposed in this paper)

表 3. 各特征维度上的详细特征提取性能(本文方法)

特征维度	精度(%)	召回率(%)	mAP (%)
釉色	88.5	86.7	85.8
纹饰	87.6	85.9	85.1
器型	92.8	91.5	90.0
工艺风格	90.2	88.3	86.5
综合	89.8	88.1	86.9

所有特征的 mAP 均在 85%以上, 综合 mAP 为 86.9%。器型特征提取精度最高(92.8%), 因其几何轮廓最为明确。纹饰特征提取精度相对最低(87.6%), 原因是部分纹饰(如缠枝纹与花鸟纹的局部片段)在视觉特征上存在一定的相似性[7] [8]。

4.5. 消融实验

本节通过逐步移除本文的关键改进模块, 定量分析各模块的独立贡献[12] [13]。实验过程中分别移除多任务共享结构(变体 A)和扩张卷积支路(变体 B), 与包含全部模块的完整模型(变体 C)进行对比, 通过综合 mAP 和推理时间的变化判断各模块对最终性能的实际贡献。消融实验设置如下。

变体 A (无多任务共享): 四个独立的单任务 ResNeXt50-SSD 模型[7] [8], 作为衡量多任务学习是否带来精度损失的参照基准。

变体 B (多任务共享, 无扩张卷积): 四任务并行输出结构, 基础特征提取网络无扩张卷积支路[9] [12], 用于单独评估多任务共享机制对推理效率的贡献。

变体 C (多任务共享 + 扩张卷积, 仅作用于纹饰支路): 即本文完整模型[13], 用于综合评估多任务共享与扩张卷积联合作用下的整体性能。

消融实验结果如表 4 所示。消融实验通过逐步移除本文方法的关键模块, 观察综合 mAP 的变化, 以

此定量判断各模块的独立贡献。变体 A 为单任务基准(无多任务共享), 变体 B 为多任务共享但无扩张卷积支路, 变体 C 为包含全部模块的完整模型。三组变体共用同一测试集, 输入输出方式一致, 唯一差异为是否包含特定模块, 因此表中的性能差异可直接归因于各模块的贡献。

Table 4. Results of ablation experiments
表 4. 消融实验结果

变体	综合 mAP (%)	推理时间(ms/张)
变体 A (无多任务共享)	—	125
变体 B (多任务共享, 无扩张卷积)	86.4	31
变体 C (本文完整模型)	86.9	33

从变体 A 到变体 B, 推理时间从 125 ms 降至 31 ms, 效率提升约 4 倍, 综合 mAP 为 86.4%, 各任务精度与单任务基准基本持平[9] [12]。从变体 B 到变体 C, 扩张卷积支路使纹饰 mAP 从 82.6%提升至 85.1%, 综合 mAP 从 86.4%提升至 86.9%, 推理时间仅增加 2 ms [13]。

消融实验表明: 多任务共享结构是效率提升的主要来源, 扩张卷积支路在几乎不增加计算成本的前提下有效改善了纹饰特征提取精度[9] [12] [13]。

5. 结论

本文针对文化创意陶瓷产品视觉特征多类别提取的需求, 提出了一种基于多任务学习的特征提取方法(MTL-CVFE), 将釉色、纹饰、器型、工艺风格四项特征提取任务整合于单一模型。主要工作与结论如下。

(1) 提出了由输入层、特征提取层和输出层构成的系统整体架构。特征提取层以 ResNeXt50-SSD 深度网络为基础, 在其上附加扩张卷积支路与四任务并行输出头, 以单次前向传播完成四类特征的提取与量化输出。

(2) 针对多任务共享基础特征提取网络的优化平衡问题、纹饰大幅面结构的感受野不足问题以及多任务特征提取的精度保障问题, 分别提出了四任务并行输出头、扩张卷积支路以及自适应加权损失函数三项技术方案。

(3) 自建了文创陶瓷产品视觉特征标注数据集, 覆盖釉色、纹饰、器型、工艺风格四类特征, 标注一致性经 Fleiss' Kappa 系数检验, 满足实验验证要求。

(4) 实验结果表明, 本文方法在自建的测试集上综合 mAP 达 85%以上, 各任务精度与单任务专用模型基本持平, 推理效率较单任务方案显著提升。消融实验验证了各改进模块的有效性。

本文方法在自建数据集上取得了良好的视觉特征提取效果, 但仍存在两方面的局限。其一, 数据集的图像来源以景德镇产区为主, 品类覆盖范围有限, 模型的泛化能力尚未在其他产区(如德化、龙泉、宜兴等)的陶瓷产品上进行验证; 不同窑口的陶瓷产品在胎质、釉色呈色、装饰技法及造型风格上均存在显著差异, 例如景德镇青花瓷的釉色呈色偏冷, 纹饰绘制精细, 而德化白瓷则以乳白釉色与刻花装饰为典型特征。这些视觉差异导致模型在跨产区数据上的特征分布发生偏移, 是提升模型泛化能力的核心科学问题之一, 亟需通过扩充多产区数据集进行跨域适应训练加以解决。其二, 模型部署依赖于 GPU 服务器, 推理效率在移动端或边缘设备上尚不具备实用性。后续研究可在上述两个方面展开: 扩充数据集的品类与产区覆盖, 评估模型的跨域泛化性能; 以及通过知识蒸馏或网络剪枝等方法压缩模型规模, 实现轻量化部署。

参考文献

- [1] 金濡欣. 陶瓷的符号与语义系统试探[J]. 中国陶瓷工业, 2019, 26(5): 14-18.
- [2] Zhao, W. (2026) Ceramic Art Design Generation and Aesthetic Quality Evaluation Based on Multimodal Large Language Models and Diffusion Probabilistic Framework. *Scientific Reports*, **16**, Article No. 26758. <https://doi.org/10.1038/s41598-026-55502-z>
- [3] Kang, X. and Zhao, Z. (2025) Optimal Design of Ceramic Form Combining Stable Diffusion Model and Gru-Attention. *Advanced Engineering Informatics*, **64**, Article ID: 103062. <https://doi.org/10.1016/j.aei.2024.103062>
- [4] 王克刚, 耿国华, 周明全. 基于纹理特征的古瓷片分类研究[J]. 计算机工程与应用, 2010, 46(12): 236-238.
- [5] Simonyan, K. and Zisserman, A. (2015) Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv: 1409.1556
- [6] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/cvpr.2016.90>
- [7] Wang, C., Dong, Q. and Lin, L. (2025) Optimisation Strategy for Ceramic Cultural and Creative Product Design Combining Analytic Hierarchy Process and Alexnet. *International Journal of Arts and Technology*, **15**, 325-342. <https://doi.org/10.1504/ijart.2025.150035>
- [8] Chen, Z., Xu, T. and Yu, P. (2024) Coupling Design Features of Material Surface Treatment for Ceramic Products Based on Resnet. *Science and Engineering of Composite Materials*, **31**, Article ID: 20240040. <https://doi.org/10.1515/secm-2024-0040>
- [9] Caruana, R. (1997) Multitask Learning. *Machine Learning*, **28**, 41-75. <https://doi.org/10.1023/a:1007379606734>
- [10] Xie, S., Girshick, R., Dollar, P., Tu, Z. and He, K. (2017) Aggregated Residual Transformations for Deep Neural Networks. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 5987-5995. <https://doi.org/10.1109/cvpr.2017.634>
- [11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., et al. (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
- [12] Ruder, S. (2017) An Overview of Multi-Task Learning in Deep Neural Networks. arXiv: 1706.05098
- [13] Yu, F. and Koltun, V. (2016) Multi-Scale Context Aggregation by Dilated Convolutions. arXiv: 1511.07122
- [14] 刘德洋, 罗中华, 沈克永, 阮英兰, 徐梦溪, 郑胜男. 一种广角-变焦协同的视觉感知装置应用设计[J]. 仪器与设备, 2026, 14(1): 19-30.
- [15] 徐梦溪, 刘德洋, 罗中华, 等. 基于抑制非高斯噪声的视觉目标跟踪方法[J]. 电声技术, 2025, 49(3): 26-28.
- [16] Liu, H. and Cosea, M. (2017) Semi-Random Partitioning of Data into Training and Test Sets in Granular Computing Context. *Granular Computing*, **2**, 357-386. <https://doi.org/10.1007/s41066-017-0049-2>