

基于改进 DenseNet 的刺绣图像分类识别的研究

刘羿漩¹, 齐振岭¹, 董苗苗², 梁允泉², 葛广英³

(1. 聊城大学 物理科学与信息工程学院 山东省光通信科学与技术重点实验室, 山东 聊城 252059;
2. 聊城大学 物理科学与信息工程学院, 山东 聊城 252059;
3. 聊城大学 计算机学院, 山东 聊城 252059)

摘要: 针对中华传统刺绣工艺传承保护问题中的分类任务, 传统的刺绣分类方法存在耗时长、精度低以及需要大量掌握专业知识的人力资源等问题; 设计了一种基于改进 DenseNet 的刺绣图像分类识别方法; 构建刺绣图像分类识别数据集; 采用局部二值模式 LBP、Canny 算子边缘提取以及 Gabor 滤波等方式提取纹理特征, 将不同特征图与原图合并为四至六通道图像数据集送入网络进行消融试验, 扩充了数据集宽度; 为稳定训练过程, 加速损失收敛速度, 提出引入 SPP (spatial pyramid pooling) 结构优化模型; 为提高分类识别精度使用 Leaky ReLU 激活函数优化 ReLU 函数; 实验结果表明基于改进 DenseNet 的刺绣图像分类识别方法可解决传统刺绣图像分类方法中存在的问题, 改进后的刺绣图像分类模型与基准模型相比准确率提高了 8.1%, 高达 97.39%。

关键词: 刺绣图像分类识别; 深度学习; 卷积神经网络; 稠密连接网络; 金字塔池化; 多通道融合

Research on Embroidery Image Classification and Recognition Based on Improved DenseNet

LIU Yixuan¹, QI Zhenling¹, DONG Miaomiao², LIANG Yunquan², GE Guangying³

(1. School of Physical Sciences and Information Engineering, Liaocheng University Shandong Provincial Key
Laboratory of Optical Communication Science and Technology, Liaocheng 252059, China;
2. School of Physical Sciences and Information Engineering, Liaocheng University, Liaocheng 252059, China;
3. School of Computer Science and Technology, Liaocheng University, Liaocheng 252059, China)

Abstract: Aiming at the categorize tasks in inheritance and protection of traditional Chinese embroidery technology, three are the problems of long time consuming, low accuracy, and massive human resources of traditional embroidery classification method, an embroidery image classification method based on improved DenseNet is proposed. The local binary pattern (LBP), Canny operator edge extraction and Gabor filtering are used to extract the texture feature and the original image, which are merged into four to six channels image data set, the data set is sent to the network to conduct the ablation test and expand the data set width. The spatial pyramid pooling (SPP) structural optimization model is proposed to accelerate the convergence rate of loss in the process of training. To improve the classification and recognition accuracy, the Leaky ReLU activation function is used to optimize the ReLU function. The simulation results show that the embroidery image classification and recognition method based on the improved DenseNet can solve the problems in the traditional embroidery image classification method. Compared with the benchmark model, the accuracy of the improved model is increased by 8.1%, which reaches up to 97.39%.

Keywords: embroidery image classification and recognition; deep learning; convolutional neural network; DenseNet; SPPnet; multi-channel fusion

0 引言

随着计算机与互联网技术的飞速发展, 计算机视觉技术的应用在日常生活中无处不在, 深度学习算法可以应用于图像分类^[1-3]、缺陷检测^[4-7]等领域。在非遗文化的数字化

保护领域, 信息技术可将非物质文化遗产转换、再现或复原成为一种可共享、可再生的数字形态, 并以全新的方式加以保存利用。韩美群等^[8]对比了近 20 年内国内外学者在非遗数字化传承方面的研究结果, 并且提出在新时代信息技

收稿日期: 2022-06-16; 修回日期: 2022-07-05。

基金项目: 中央引导地方科技发展专项基金(YDZX2017370000283)。

作者简介: 刘羿漩(1998-), 女, 黑龙江青冈人, 硕士生, 主要从事图像处理与模式识别方向的研究。

通讯作者: 葛广英(1964-), 男, 山东聊城人, 博士, 教授, 硕士生导师, 主要从事计算机视觉、模式识别和机器视觉方向的研究。

引用格式: 刘羿漩, 齐振岭, 董苗苗, 等. 基于改进 DenseNet 的刺绣图像分类识别的研究[J]. 计算机测量与控制, 2023, 31(1): 194-201.

术高速发展的阶段应当增强数字技术在高质量文化供给上的支撑作用。在中国众多非遗文化中, 中华传统刺绣可按地区、工艺以及加工特点等标准分为不同种类, 进而使用在不同的场合, 而区分不同刺绣种类的过程通常相当繁琐复杂且需要大量掌握相关专业知识的人力成本与时间成本, 所以可以运用深度学习算法实现刺绣的批量分类。杨蕾等^[9]提出一种能够将绣品针法纹样特征进行提取和匹配的方法, 构建了一种以导向滤波作为尺度变换函数的特征提取模型, 匹配准确率高达 88%。周泽聿等^[10]通过对 Xception 模型全连接层的层数和 dropout 等参数进行微调并探讨了不同参数对结构的影响, 提出了一种基于 Xception-TD 的中华传统刺绣分类模型, 该模型的准确率高达 96.86%, 且在刺绣多分类的问题上准确率也优于基准模型。

图像处理是深度学习算法最早尝试应用的领域, 其中卷积神经网络 (convolutional neural network, CNN)^[11]是深度学习的代表算法之一。1989 年加拿大多伦多大学教授 Yann LeCun 提出了首个卷积神经网络 LeNet。后来随着硬件水平提升。2012 年 10 月, Hinton 教授以及他的学生采用更深的卷积神经网络模型 AlexNet 在著名的大规模数据集视觉识别比赛 (ImageNet large scale visual recognition challenge, ILSVRC) 上取得了第一名且效果远超第二名, AlexNet 网络的高性能使得图像识别的领域研究更进一步。随后, VGGNet 在 2014 年的 ILSVRC 比赛中, 赢得了定位任务的冠军, 在分类任务中排名第二。但 AlexNet、VGGNet 等模型主要通过增加卷积神经网络的深度以达到更好的分类效果, 网络层数增加的同时也带来了梯度消失、过拟合、梯度爆炸等问题。2014 年面世的 GoogleNet 网络使用 1×1 卷积核进行降维以及映射处理, 引入了 Inception 结构以融合不同尺度特征信息^[12]。在 2015 年的 ILSVRC 比赛中, ResNet 利用一种称为 Shortcut Connection 结构来跳过网络的某些层, 以 3.57% 的 top5 错误率赢得了冠军^[13]。基于 ResNet 模型训练生成的网络中存在层贡献很少的局限性, Liu 等于 2017 年提出了 DenseNet, 使用了跨层连接及以前馈方式, 每个层都会接受前面所有层作为其额外的输入^[14]。相较于 ResNet, DenseNet 旁路加强了特征的重用且具有更少的参数量。Google 公司在 2017 年提出了 MobileNet, 它是一种小巧而高效的 CNN 模型, 优点在于提出了深度可分离卷积, 降低了计算量和参数量^[15], 使卷积神经网络更易部署到移动端。2020 年, Google 公司提出了 EfficientNet 模型^[16], 该模型是一个快速高精度模型, 它使用了深度 (depth)、宽度 (width)、输入图片分辨率 (resolution) 共同调节技术。卷积神经网络的优势在于无需进行人为的特征提取, 近些年卷积神经网络推动了很多行业的发展, 本文将卷积神经网络算法应用到刺绣分类问题上并对比分析结果。

本文对比了 Googlenet、Resnet、DenseNet、MobileNet 以及 EfficientNet 共 5 种常用的经典 CNN 模型在中华传统

刺绣分类数据集上的表现后, 选择基于 DenseNet 模型进行改进, 采用 SPP (spatial pyramid pooling) 模块^[19]加快损失收敛, 稳定算法训练过程, 通过修改激活函数提高模型精度, 然后采用局部二值模式 LBP^[17-18]、Canny 算子边缘提取以及 Gabor 滤波等方式提取纹理特征, 与原图合并成四至六通道图像作为卷积神经网络的输入进行训练, 扩充数据宽度, 便于算法提取特征, 进一步提升刺绣分类的识别准确率。

1 基于改进 DenseNet 模型分类方法

1.1 DenseNet 模型

DenseNet 卷积神经网络模型是由 Huang 等在 2017 年提出^[14], 该网络借鉴了 ResNet 结构, 即建立前面层与后面层的连接, 利于梯度反向传播和训练更深层次网络。DenseNet 模型能够实现特征重复使用, 减轻了梯度消失问题, 减少参数量并提高特征的传播率和利用率。DenseNet 卷积神经网络模型中的密集连接块 DenseBlock 比 ResNet 的连接机制更为激进, 所有层之间互相连接, 每个层的额外输入包括了前面的所有层, 在 DenseBlock 内特征图大小统一, 所有层互相连接以使每一层能够接受前面所有层中的特征, 加强特征重用。

假设 $H(\cdot)$ “表示卷积神经网络的非线性效应, 其中包括卷积、池化以及激活等多种操作的组合。传统 CNN 网络经过第 i 层后得到的输出 x_i 的计算方式如式 (1) 所示:

$$x_i = H_i(x_{i-1}) \quad (1)$$

对 ResNet 卷积神经网络模型来说, 第 i 层的输出 x_i 的计算方式如式 (2) 所示:

$$x_i = H_i(x_{i-1}) + x_{i-1} \quad (2)$$

对 DenseNet 卷积神经网络模型来说, 第 i 层的输出 x_i 将连接第 i 层之前所有的层作为输入, 则第 i 层的输出 x_i 的计算方式如式 (3) 所示:

$$x_i = H_i([x_{i-1}, \dots, x_2, x_1, x_0]) \quad (3)$$

式 (3) 中的非线性效应 $H_i(\cdot)$ 为归一化 (batch normalization, BN) 操作。DenseBlock 结构连接方式如图 1 所示。



图 1 DenseBlock 结构图

由图 1 可知 DenseNet 模型的前向传播方式是比 ResNet 模型更为密集的连接, 各个层的特征图在通道维度上连接在一起使整个网络更易训练。

1.2 改进 DenseNet 模型

1.2.1 SPP 结构

为解决图像分类任务中输入图像尺寸限制问题, 2015

年, He 等^[20] 提出金字塔池化 (spatial pyramid pooling, SPP) 结构, 若要使卷积神经网络适应不同尺寸的输入图片, 文献 [20] 提出了在卷积神经网络结构后面引入 SPP 结构。

具体操作为最后一层卷积层和全连接层前加入 SPP 结构, 对通过最后一层卷积的多层特征图进行通道调整, 本文将每个特征图按 16×16 , 4×4 以及 1×1 不同大小划分为不同尺寸的网格, 然后对每个网格进行最大池化 (Max pooling) 操作, 然后经过拼接通道将特征图转化为固定长度的特征向量, 使其作为后面全连接层的输入。总体来说, SPP 结构就是通过不同大小窗口进行池化, 得到分辨率不同的特征图, 对每个特征图连接 1×1 卷积层来压缩调整通道维度。SPP 结构图如图 2 所示。

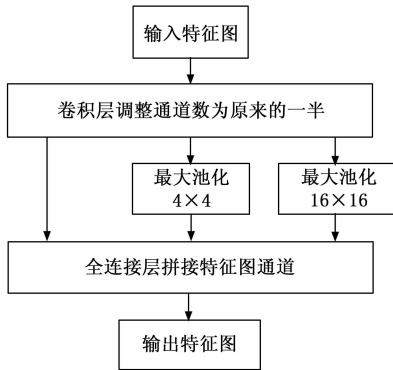


图 2 SPP 结构图

1.2.2 激活函数

卷积神经网络中激活函数用于实现网络的非线性映射, DenseNet 网络结构中使用的激活函数为 ReLU 函数, 但 ReLU 函数在输入小于 0 时无法激活神经元, 易使网络训练不收敛, ReLU 函数定义如式 (4) 所示。为解决此问题, Leaky ReLU 函数^[21]利用较小的权值保留了部分信息, 其定义如式 (5) 所示:

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (4)$$

$$g(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases} \quad (5)$$

本文将 Leaky Relu 函数中系数 α 取值为 0.01, 有助于扩大 ReLU 函数的范围, 该激活函数将激活函数的整个函数范围定义在负无穷到正无穷上, 使 x 的较小的线性分量作为负输入调整网络中负值的零梯度问题。实验比较了激活函数选择 ReLU 函数以及 Leaky ReLU 函数对刺绣图像分类模型性能的影响, 实验证明了激活函数选为 Leaky ReLU 函数能够提升网络的性能。

引入 SPP 结构以及进行激活函数改进后的 DenseNet 整体网络结构如图 3 所示。

其中 DenseBlock 中的激活函数全部为 Leaky ReLU 函数。

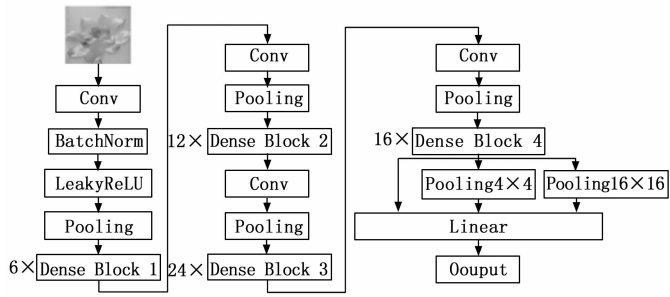


图 3 改进后的 DenseNet 网络结构图

1.2.3 数据集宽度扩充

1) 局部二值模式 (local binary pattern, LBP) 是一种用来描述图像局部纹理特征的算子, 提取出的特征是图像的局部纹理特征, 具有旋转不变性和灰度不变性等优点。具体的计算方式如式 (7) 所示:

$$IMG_{lbp} = \sum_{i=1}^8 2^{i-1} \times F(x_0, x_i) \quad (7)$$

$$F(x_0, x_i) = \begin{cases} 0 & x_i < x_0 \\ 1 & x_i \geq x_0 \end{cases}$$

式 (7) 中 x_0 表示一个小窗口的中心点像素灰度值, x_i 表示中心点周围的其他像素点的灰度值, 共有 8 个像素点, $F(\cdot)$ 表示将小窗口中心点像素与相邻像素比较得到的二值序列, 具体方法为将中心像素点的值与周围像素点的值做对比, 若中心像素点的值小于周围像素点的值, 则对应周围像素点标记为 1, 反之, 若中心像素点的值大于周围像素点的值, 则将对应像素标记为 0, 将得到的二值序列与二值化后的周围像素点对应相乘后并相加, 即得到中心像素值。

根据刺绣图像分类数据集中的 RGB 图像利用 LBP 特征提取前后对比如图 4 所示。

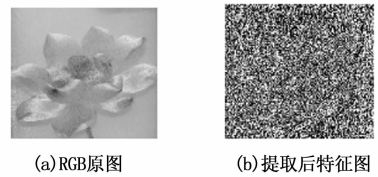


图 4 LBP 提取特征前后对比图

2) 基于 Canny 算子的边缘检测可以通过高斯滤波、像素梯度计算、非极大值抑制等步骤简化图像, 改变图像的表现形式使其更易提取特征。

首先对图像进行高斯滤波, 即将一个灰度图像与一个高斯核进行卷积操作, 具体计算方式如式 (8) 所示:

$$IMG'_{(x,y)} = G * IMG_{(x,y)} \quad (8)$$

其中: $IMG'_{(x,y)}$ 为原图经过高斯滤波后图像中在 (x, y) 位置像素点的灰度值, $IMG_{(x,y)}$ 为 RGB 原图转变后的灰度图在 (x, y) 位置像素点的灰度值, G 表示标准差为 σ 的二维高斯卷积核, 其定义如式 (9) 所示:

$$G = \frac{1}{2\pi\sigma} e^{-x^2+y^2/2\sigma^2} \quad (9)$$

然后计算经过高斯滤波后的图像的梯度值与梯度方向, 以此观察灰度值变化最大的像素点, 梯度值与梯度方向的具体计算方式如式 (10) 和式 (11) 所示:

$$\nabla f(m,n) = \sqrt{G_x^2 + G_y^2} \quad (10)$$

$$\theta = \arctan \frac{G_y}{G_x} \quad (11)$$

其中: G_x 为像素点在 x 方向上的梯度值, G_y 为像素点在 y 方向上的梯度值。然后进行非极大值抑制不是极大值的像素点, 过滤非边缘像素点, 使边缘宽度尽可能小, 接着进行双阈值检测设置一个低阈值和一个高阈值, 进一步减少边缘像素点。

根据刺绣图像分类数据集中的 RGB 图像在经过 Canny 算子提取边缘的前后对比如图 5 所示。

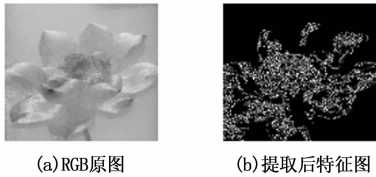


图 5 Canny 算子边缘检测前后对比图

3) Gabor 滤波器是边缘提取的线性滤波器, 其频率和方向表达与人类视觉系统类似, 能够提供良好的方向选择和尺度选择特性, 而且对于光照变化不敏感, 因此十分适合纹理分析。它的优势体现在 Gabor 滤波器可以从不同方向和尺度上对物体纹理进行提取特征, 并进行随意组合。本文采取二维 Gabor 滤波器对图像进行滤波, 其滤波核可以分为实部和虚部, 其具体数学表达式如式 (12) 所示:

$$\begin{cases} g_{\text{real}}(x,y;\lambda,\theta,\psi,\sigma,\gamma) = e^{-x'^2+r'y'^2/2\sigma^2} \cos(2\pi \frac{x'}{\lambda} + \psi) \\ g_{\text{imag}}(x,y;\lambda,\theta,\psi,\sigma,\gamma) = e^{-x'^2+r'y'^2/2\sigma^2} \sin(2\pi \frac{x'}{\lambda} + \psi) \end{cases} \quad (12)$$

其中: x' 与 y' 的计算方式如式 (13) 所示:

$$\begin{cases} x' = x \cos \theta + y \sin \theta \\ y' = -x \sin \theta + y \cos \theta \end{cases} \quad (13)$$

上面两个公式中, 参数 λ 为波长, 可影响滤波器滤波尺度; 参数 θ 为滤波器方向, 共有 4 个, 分别为 0 度, 45 度, 90 度以及 135 度; 参数 ψ 为相位偏移, 取值范围为 -180 至 180; 参数 γ 为空间纵横比, 本文取值为 0.5; 参数 σ 为带宽, 本文取值为 2π 。

对刺绣图像分类数据集中的 RGB 图像进行二维 Gabor 滤波前后对比如图 6 所示。

1.3 改进 DenseNet 模型训练流程

本文在已有研究基础上对图像分类网络进行研究改进, 根据训练样本图像分类数据集 $X \{x_1, x_2, \dots, x_n\}$ 以及对应的 m 种类别 $Y \{y_1, y_2, \dots, y_m\}$, 对网络结构 $f(\theta)$, 损失函数 l 以及学习率 L_R 进行训练, 算法训练流程如图 7

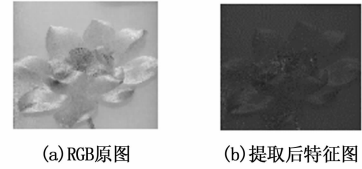


图 6 Gabor 滤波前后对比图

所示。

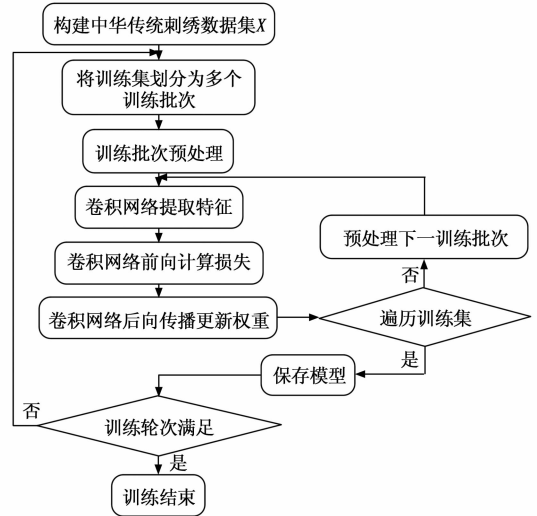


图 7 刺绣分类的训练过程

1) 按预先设定的 Batch size 从训练集 $X \{x_1, x_2, \dots, x_n\}$ 中选取一个训练批次 $X_1 \{x_a, \dots, x_b\}$ 进行归一化预处理, 对于训练批次中每张刺绣图像中 RGB 三通道分别按均值 $mean_{r,g,b} = [0.485, 0.456, 0.406]$, 标准差 $std_{r,g,b} = [0.229, 0.224, 0.225]$ 进行如式 (14) 操作:

$$\begin{aligned} Output_r &= \frac{Input_r - mean_r}{std_r} \\ Output_g &= \frac{Input_g - mean_g}{std_g} \\ Output_b &= \frac{Input_b - mean_b}{std_b} \end{aligned} \quad (14)$$

式 (1) 中 $Input_r$ 、 $Input_g$ 、 $Input_b$ 为未归一化前图像的 RGB 三通道像素值, $Output_r$ 、 $Output_g$ 、 $Output_b$ 为归一化后图像的 RGB 三通道像素值。

2) 将归一化后的训练批次 $X'_1 \{x'_a, \dots, x'_b\}$ 送入网络进行前向传播提取特征, 在 Denseblock 中第 $n-1$ 之前所有层中特征图互相连接并卷积得到第 n 层所得特征图, 计算公式如式 (15) 所示。

$$M_n = F(M_1 \oplus M_2 \oplus \dots \oplus M_{n-1}) \quad (15)$$

式 (2) 中 M_n 表示网络第 n 层输出特征, F 表示卷积操作, \oplus 表示特征之间相互连接操作, 由此可得第 n 层网络有个 $\frac{n^2+n}{2}$ 连接操作。

3) 训练批次经网络训练后得到对应类别 Y'_1 , 再根据

预测类别 Y'_1 与真实类别 Y_1 计算损失 l_1 ，如式 (16) 所示：

$$Y'_1 = f(\theta, X'_1) \tag{16}$$

$$l_1 = L(Y_1, Y'_1)$$

式 (3) 中利用损失函数 L 计算损失 l ，训练过程的损失函数选用交叉熵损失函数，如式 (17) 所示：

$$l_1 = -\frac{1}{b-a} \sum_{i=1}^{b-a} \sum_{c=1}^m p_k \log(q_k) \tag{17}$$

式 (17) 中 $b-a$ 表示每一训练批次中的样本个数， c 表示类别数量共有 m 种， p_k 为符号函数，若训练批次中第 i 个样本的真实类别与类别 c 相同则取 1，反之取 0， q_k 表示预测训练批次中第 i 个样本属于类别 c 的概率。

4) 计算得到的损失 l 后向传播更新网络权重 θ ，如式 (18) 所示：

$$\theta' = \theta - l_r \times \nabla g(l_1) \tag{18}$$

5) 判断整个数据集是否遍历结束，若遍历未结束，选取下一训练批次重复以上步骤，若遍历结束，则保存该轮次的模型权重参数文件并判断是否满足训练轮次。

6) 若未满足训练轮次，重复以上步骤，若满足训练轮次则结束模型训练。

1.4 改进 DenseNet 模型测试流程

本次实验需要对训练所得的神经网络模型进行测试以验证效果，主要采用的评估指标为准确率 (accuracy)、精确率 (precision)、召回率 (recall) 和特异度 (specificity) 以及混淆矩阵。刺绣图像分类模型的测试流程如图 8 所示。

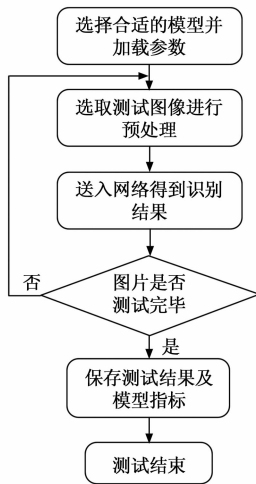


图 8 刺绣分类的测试过程

具体的测试步骤如下：

- 1) 选择训练得到的卷积神经网络模型并加载；
- 2) 从刺绣图像分类数据集中选取待测刺绣图像，进行归一化预处理；
- 3) 将预处理后的刺绣图像送入网络中进行计算并得到识别分类结果；
- 4) 判断测试集中图片是否全部测试完毕，若未完毕，重复以上步骤，若完毕，保存测试结果并计算模型的评估

指标。

2 实验与结果分析

2.1 数据集采集与扩充

本文数据集样本采集来源于网络，筛选出八类中华传统刺绣，包括中国四大名绣：苏绣、湘绣、粤绣、蜀绣以及少数民族刺绣：哈密刺绣、水族马尾绣、西秦刺绣和其它类别，其中其它类别内的刺绣还包含京绣、鲁绣、盘金绣以及十字绣等。数据集中的部分刺绣图像（随机选取一张）如图 9 所示。整个数据集按照 9：1 的比例划分为训练集与验证集，即训练集图像共 6 933 张，测试集图像共 767 张。具体类别分布如图 10 所示。



图 9 刺绣图像分类数据集样本

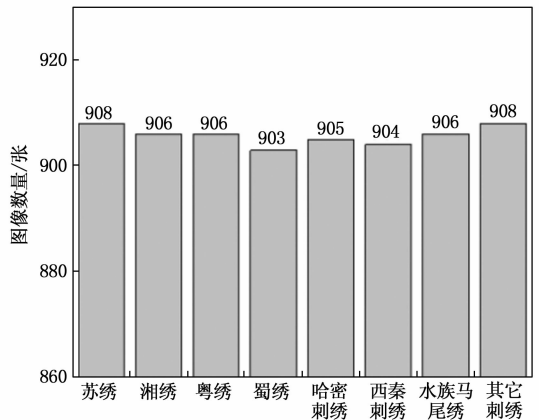


图 10 刺绣图像分类数据集类别分布图

2.2 实验环境及参数

本文实验的训练与测试环境相同，均为 Windows10 操作系统，使用 Pytorch 深度学习框架来实现整个模型训练与测试过程，实验环境的具体参数如表 1 所示。

表 1 基准模型对比仿真环境

仿真环境	配置
CPU	Intel(R) Xeon(R) Silver 4210R CPU @2.40 GHz
GPU	NVIDIA Quadro RTX 4 000
内存	64 G
Python 版本	3.7
Pytorch 版本	1.10.0

实验选取 5 个常用的经典卷积神经网络模型进行基准

模型对比实验, 分别为 GoogleNet (Inception V3)、ResNet50、DenseNet-121、MobileNetV2 与 EfficientNet-B0, 然后将这 5 种模型与本文改进后的模型进行性能对比, 这 5 种模型的详细参数对比分析如表 2 所示。

表 2 卷积神经网络模型对比表

网络模型	年份	Accuracy(top-1)	Accuracy(top-5)	参数量
InceptionV3	2015	77.9%	93.7%	23.9M
ResNet50	2015	74.9%	92.1%	25.6M
DenseNet121	2017	75.0%	92.3%	8.1M
MobileNetV2	2018	71.3%	90.1%	3.5M
EfficientNetB0	2020	76.3%	93.2%	1.7M

经过反复尝试, 本次实验中选定的训练相关超参数值如表 3 所示。

表 3 网络训练超参数

超参数	取值
类别数目 Num class	8
优化器 Optimizer	Adam
学习率 learning rate	1E-04
训练轮次 Num epochs	200
批大小 Batch size	16
预训练 Pre-training	No

2.3 实验及分析

2.3.1 不同算法实验对比分析

本次实验需要对训练所得的神经网络模型验证效果, 主要采用的评估指标为准确率 (accuracy)、精确率 (precision)、召回率 (recall) 和特异度 (specificity) 以及混淆矩阵。

1) 混淆矩阵。

混淆矩阵是评判模型优劣的一种指标, 常用于评判分类器模型的优劣。表 4 代表一个二分类示例的混淆矩阵。

表 4 二分类混淆矩阵

混淆矩阵		真实值	
		Positive	Negative
预测值	Positive	TP	FP
	Negative	FN	TN

2) 准确率 Accuracy。

准确率 (ACC) 表示模型正确分类样本占总样本数比例 (包括所有类别)。计算公式如下:

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

3) 精确率 Precision。

精确率 (PPV) 表示模型预测的所有 Positive 中预测正确的比例。计算公式如下:

$$Precision = \frac{TP}{TP + FP}$$

4) 召回率 Recall。

召回率/灵敏度 (TPR) 表示在所有真实 Positive 中,

模型预测正确的 Positive 比例。计算公式如下:

$$Recall = \frac{TP}{TP + FN}$$

5) 特异度 Specificity。

特异度 (TNR) 表示在所有真实 Negative 中, 模型预测正确的 Negative 比例。计算公式如下:

$$Specificity = \frac{TN}{TN + FP}$$

GoogleNet、ResNet、DenseNet、MobileNet 与 EfficientNet 模型在各个种类上取得的精确率、召回率和特异度如表 5 所示。

表 5 基准模型测试结果

	苏绣			湘绣		
	pre	rec	spe	pre	rec	spe
GoogleNet	0.41	0.631	0.834	0.445	0.453	0.898
ResNet	0.61	0.590	0.929	0.632	0.632	0.934
DenseNet	0.851	0.984	0.967	0.963	0.880	0.994
MobileNet	0.655	0.795	0.921	0.772	0.607	0.968
EfficientNet	0.837	0.844	0.969	0.850	0.726	0.977
	粤绣			蜀绣		
	pre	rec	spe	pre	rec	spe
GoogleNet	0.628	0.639	0.628	0.639	0.628	0.639
ResNet	0.569	0.689	0.569	0.689	0.569	0.689
DenseNet	0.951	0.824	0.951	0.824	0.951	0.824
MobileNet	0.713	0.689	0.713	0.689	0.713	0.689
EfficientNet	0.740	0.815	0.740	0.815	0.740	0.815
	哈密刺绣			西秦刺绣		
	pre	rec	spe	pre	pre	rec
GoogleNet	0.976	0.909	0.999	0.938	0.976	0.909
ResNet	1.000	0.864	1.000	0.961	1.000	0.864
DenseNet	0.911	0.932	0.994	0.962	0.911	0.932
MobileNet	0.955	0.955	0.997	0.912	0.955	0.955
EfficientNet	0.953	0.932	0.997	1.000	0.953	0.932
	水族马尾绣			其它刺绣		
	pre	rec	spe	pre	pre	rec
GoogleNet	0.949	0.989	0.949	0.989	0.949	0.989
ResNet	0.931	1.000	0.931	1.000	0.931	1.000
DenseNet	0.989	0.989	0.989	0.989	0.989	0.989
MobileNet	0.958	0.979	0.958	0.979	0.958	0.979
EfficientNet	0.969	0.989	0.969	0.989	0.969	0.989

由表 5 可知, 基于 DenseNet 的刺绣图像分类模型在 8 个刺绣种类上取得的精确率 (precision)、召回率 (recall) 和特异度 (specificity) 与其他卷积神经网络模型相比较且分布更均匀, 证明了 DenseNet 模型较其他卷积网络模型更适用于刺绣图像分类任务。基于 DenseNet 的刺绣图像分类模型在刺绣图像分类数据集上获得的混淆矩阵如图 11 所示。

从图 11 中可以看出在刺绣分类数据集中基于 DenseNet 的中华传统刺绣分类模型在水族马尾绣种类上识别率最高, 而在粤绣种类上准确率最低。由混淆矩阵可知各类别识别正确概率均在 80% 以上, 说明基于 DenseNet 的刺绣图像分

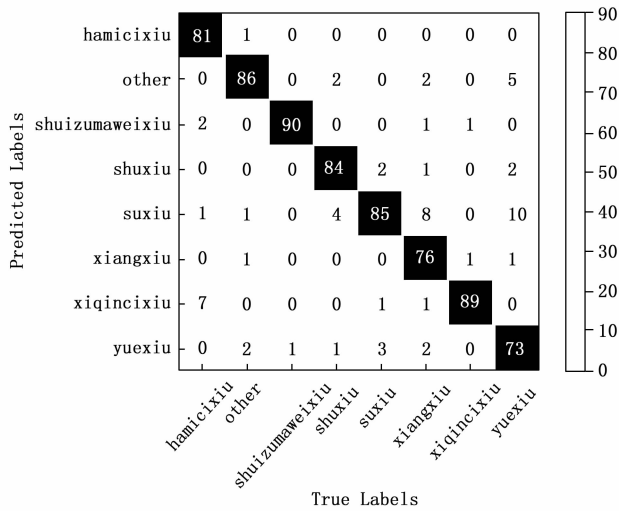


图 11 刺绣图像分类数据集中八种刺绣的混淆矩阵

类模型对数据集中存在的噪声和带有遮挡与具有相似特征的刺绣图像，有较强的鲁棒性。

因此选择基于 DenseNet 网络结构进行改进优化，在模型中引入 SPP 结构并进行激活函数优化，改进后的算法与其他算法在训练过程中的损失收敛对比过程如图 12 所示。

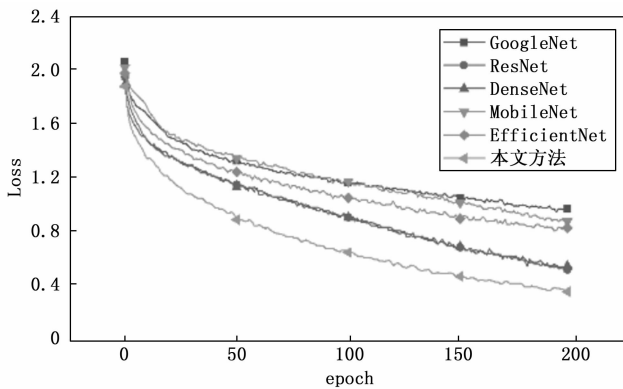


图 12 各算法损失收敛对比图

由图 12 可得本文方法在刺绣图像分类识别数据集上的损失收敛速度比其他算法的损失收敛速度更快。

为验证改进后的算法的优越性有效性，在刺绣图像测试集上与其它算法进行对比，各算法取得的准确率如表 6 所示，利用本文方法改进后的算法在刺绣图像测试集上取得的准确率远远高于其他基准卷积神经网络算法和改进前的 DenseNet 算法，进一步证明了本文提出方法的有效性。

表 6 算法取得的准确率对比

模型	准确率(ACC)/%
GoogleNet	64.02
ResNet	86.22
DenseNet	89.29
MobileNet	72.63
EfficientNet	80.47
本文方法	92.39

2.3.2 数据集宽度扩充消融实验

通过组合刺绣图像分类数据集中的 RGB 三通道原图及以上所介绍的 4 种方法得到的特征图，分别组合得到 8 种不同方案，分别为 RGB、RGB+LBP、RGB+Canny、RGB+Gabor、RGB+LBP+Canny、RGB+LBP+Gabor、RGB+Canny+Gabor 以及 RGB+LBP+Canny+Gabor。从准确率与预测时间 2 个方面对比数据集宽度扩充前后基于改进 DenseNet 的刺绣图像分类模型性能的差异，如表 7 所示。

表 7 数据集宽度扩充消融试验结果分析

Schemes	Channels	Prediction Time/s	Accuracy /%
RGB	3	0.624	92.39
RGB+LBP	4	0.727	97.17
RGB+Canny	4	0.732	97.23
RGB+Gabor	4	0.805	97.39
RGB+LBP+Canny	5	0.826	92.03
RGB+LBP+Gabor	5	0.749	93.01
RGB+Canny+Gabor	5	0.773	90.53
RGB+LBP+Canny+Gabor	6	0.954	87.43

将数据集按照 8 种不同宽度扩充方案送入改进后的刺绣图像分类模型进行训练，经过 Savitzky-Golay 滤波平滑处理后，得到 8 种不同方案的分类模型的损失收敛对比图如图 13 所示，在刺绣图像分类数据集上的取得的准确率对比图如图 14 所示。

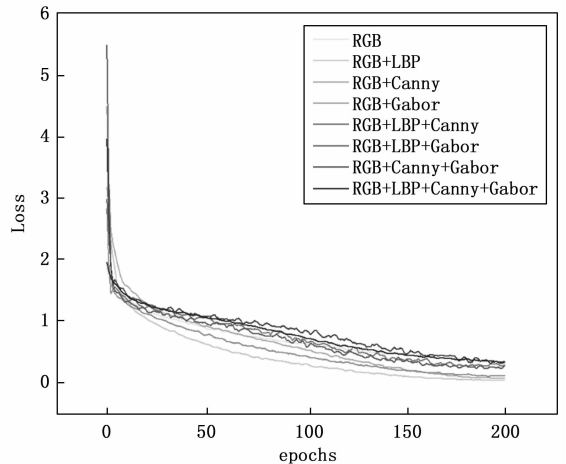


图 13 数据集宽度扩充消融试验损失收敛对比图

在通道数增加至四通道时，刺绣图像分类模型的准确率明显提高，达到了 97.39%，但方案 4 预测时间比方案 2 多了 0.1 s，准确率并未显著提升。综合预测时间与模型在刺绣图像分类测试集上达到的准确率两方面综合来看应选择方案 2，即将 RGB 图像与 LBP 提取的特征图合并为四通道图像数据作为网络输入。

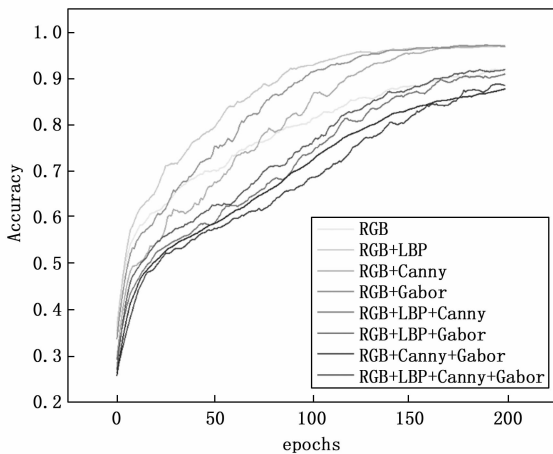


图 14 数据集宽度扩充消融试验准确率对比图

3 结束语

为提升刺绣分类效果, 构造了刺绣图像分类数据集, 研究了基于卷积神经网络的刺绣图像分类模型, 并基于 DenseNet 模型提出引入 SPP 算法以及替换激活函数优化网络结构, 提出不同扩充数据集宽度方法并进行消融实验, 确定最佳输入网络数据集应为 RGB 图像与 LBP 提取的特征图合并的四通道图像数据, 改进后的刺绣图像分类模型与基准模型相比, 准确率提高了 8.1%, 高达 97.39%, 平均分类识别时间为 72.7 ms, 验证了改进后算法具有较高的识别精度与较快识别速度, 提升了刺绣图像分类效果。

本文进行的刺绣图像分类数据集类别有限且大部分来源于网络, 未来可进一步收集更丰富的刺绣类别对模型进行训练, 提升识别分类效果。

参考文献:

- [1] 宋雯琦, 闫德勤, 刘德山, 等. 基于空谱特征的核极端学习机高光谱遥感图像分类算法 [J]. 计算机测量与控制, 2020, 66 (11): 175-181.
- [2] 曹辉, 翁理国, 张德正. 基于密集连接空洞卷积神经网络的青藏地区云雪图像分类 [J]. 计算机测量与控制, 2019, 252 (9): 169-173.
- [3] 谢亮. 弱光照环境下人脸表情图像识别方法研究 [J]. 计算机测量与控制, 2017, 228 (9): 234-236, 241.
- [4] 张广世, 葛广英, 朱荣华, 等. 基于改进 YOLOv3 网络的齿轮缺陷检测 [J]. 激光与光电子学进展, 2020, 57 (12): 153-161.
- [5] 李泽辉, 陈新度, 黄佳生, 等. 基于改进 YOLOv3 的纹理瓷砖缺陷检测 [J]. 激光与光电子学进展, 2022, 59 (10): 294-302.
- [6] 王振亚, 赵继红, 王艳鹏, 等. 基于改进 EfficientNet 网络的药片检测分类的研究 [J]. 现代计算机, 2021, 27 (28): 27-32.
- [7] 李可, 祁阳, 宿磊, 等. 基于改进 ACGAN 的钢表面缺陷视觉检测方法 [J/OL]. 机械工程学报: 1-9. [2022-06-

- 15]. <http://kns.cnki.net/kcms/detail/11.2187.TH.20220526.1827.106.html>
- [8] 韩美群, 周小芹. 近二十年来非物质文化遗产数字化传承研究回顾与展望 [J]. 中南民族大学学报 (人文社会科学版), 2022, 42 (1): 65-74, 184.
- [9] 杨蕾, 胡慧, 周军. 刺绣针法图样特征点提取及匹配方法研究 [J]. 计算机应用研究, 2021, 38 (7): 2231-2234, 2240.
- [10] 周泽聿, 王昊, 张小琴, 等. 基于 Xception-TD 的中华传统刺绣分类模型构建 [J]. 数据分析与知识发现, 6 (21): 338-347.
- [11] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN Features off-the-shelf: an Astounding Baseline for Recognition [C] // 2014 IEEE conference on computer vision and pattern recognition workshops. IEEE, 2014.
- [12] SZEGEDY C, LIU W, JIA Y, et al. Going deeper with convolutions [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1-9.
- [13] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C] // Computer Vision and Pattern Recognition, 2016: 770-778.
- [14] HUANG G, LIU Z, DER MAATEN L V, et al. Densely Connected Convolutional Networks [C] // Computer Vision and Pattern Recognition, 2017: 2261-2269.
- [15] HOWARD A G, ZHU M, CHEN B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications [Z]. 2017, 4.
- [16] TAN M X, LE Q V. EfficientNet: rethinking model scaling for convolutional neural networks [C] // Proceedings of the 36th International Conference on Machine Learning. Long Beach, CA: PMLR, 2019: 6105-6114.
- [17] HEIKKILA M, PIETIKAINEN M. A texture-based method for modeling the background and detecting moving objects [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28 (4): 657-662.
- [18] DING Y, LI W H, FAN J T, et al. Robust moving object detection under complex background [J]. Computer Science and Information Systems, 2010, 7 (1): 201-210.
- [19] HARALICK R M, SHANMUGAM K, DINSTEN I. Textural Features for Image Classification [J]. IEEE Transactions on Systems Man & Cybernetics, 1973, SMC-3 (6): 610-621.
- [20] HE K, ZHANG X, REN S, et al. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37 (9): 1904-1916.
- [21] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C] // Proc of International Conference on Machine Learning. Cambridge: MIT Press, 2013: 30 (1): 3.