

文章编号: 2095-2163(2022)01-0146-09

中图分类号: TP391.1

文献标志码: A

# 基于融合特征和迁移学习的艺术图像情感识别研究

杨松<sup>1,2,3</sup>, 刘佳欣<sup>1</sup>, 潘建达<sup>1</sup>

(1 大连外国语大学 软件学院, 辽宁 大连 116044; 2 大连外国语大学 语言智能中心, 辽宁 大连 116044;

3 网络空间多语言大数据智能分析研究中心, 辽宁 大连 116044)

**摘要:** 艺术图像的视觉情感分析有助于对艺术图像的鉴赏与保护。本文利用改进的 FPN 模型提取艺术图像的不同层次的情感语义特征, 运用 CLAHE 突出艺术图像的颜色特征, 分别提取图像在 HSV 颜色空间下的 H-S 二维特征及在 YCrCb 颜色空间下的 CrCb 二维特征, 并设置两种颜色特征的不同权重, 提出一种基于图像特征融合的情感识别模型 HCFNet。实验结果表明, 本文提出的模型对艺术图像情感识别的准确率可达到 90.63%, 比经典卷积神经网络模型和未改进 FPN 模型均有提升, 可有效实现艺术图像情感识别任务。

**关键词:** 艺术图像; 视觉情感分析; 颜色特征; 特征融合

## Research on emotion recognition of artistic images based on fusion features and transfer learning

YANG Song<sup>1,2,3</sup>, LIU Jiixin<sup>1</sup>, PAN Jianda<sup>1</sup>

(1 School of Software, Dalian University of Foreign Languages, Dalian Liaoning 116044, China;

2 Language Intelligence Center, Dalian University of Foreign Languages, Dalian Liaoning 116044, China;

3 Research Center for Networks Space Multi-Languages Big Data Intelligence Analysis, Dalian Liaoning 116044, China)

**[Abstract]** The visual emotion analysis of artistic images is helpful to the appreciation and protection of artistic images. The improved FPN model is used to extract different levels of emotional semantic features of art images, and CLAHE is used to highlight the color features of art images. The H-S two-dimensional features of images in HSV color space and the crcb two-dimensional features in YCrCb color space are extracted respectively, and the different weights of the two-color features are set, an emotion recognition model hcfnet based on image fusion is proposed. The experimental results show that the accuracy of the proposed model for art image emotion recognition can reach 90.63%, which is improved compared with the classical convolutional neural network model and the unmodified FPN model, and can effectively complete the task of art image emotion recognition.

**[Key words]** artistic image; visual sentiment analysis; color features; features fusion

## 0 引言

随着图像情感计算的深入推进, 研究方法从传统的手工提取特征过渡到表示学习阶段, 深度学习模型的运用能够有效弥合图像底层视觉特征与深层语义特征之间的“语义鸿沟”<sup>[1]</sup>。李志义等对比传统图像特征提取方法和深度学习方法后, 利用改进的 VGGNet 模型, 将增强颜色特征和增强纹理特征的图像分别作为模型输入, 以提升图像情感识别的准确率<sup>[2]</sup>; 杨子文等提出一种基于两层迁移卷积神经网络的图像情感识别方法, 采用 ImageNet2012 数据集提取图像普适的底层视觉特征, 将底层特征权重迁移至模型结构相同的网络中, 利用“Painters by numbers”抽象风格分类数据集微调网络模型来提取

图像深层语义特征<sup>[3]</sup>; 蓝亦伦等利用自动编码器构建视觉特征和语义特征联合嵌入网络, 缩小两者之间的语义鸿沟, 提取图像显著区域特征, 引入注意力机制建立显著区域与视觉语义联合嵌入特征两者之间的关联, 确定显著情感区域, 并在此基础上构建图像情感分类器, 实现图像情感分类<sup>[4]</sup>。为解决缺失情感标签的图像, 降低卷积神经网络模型训练性能问题, 杨文武等构建大型数据集, 并将图像主体与背景分离, 在 CNN 基础上利用关系学习网络模型提取图像层级特征信息及关系, 以弥补图像特征与情感语义之间存在的鸿沟, 进而精准实现图像情感分类任务<sup>[5]</sup>。

目前人脸表情识别、风景及场景图像等成为图像情感识别领域的研究热点, 而以颜色搭配、线条手

**基金项目:** 国家自然科学基金(61806038); 辽宁省社会科学规划基金(L18BTDQ005); 辽宁省教育厅科学研究项目(2019JYT07)。

**作者简介:** 杨松(1977-), 男, 博士, 副教授, CCF 会员, 主要研究方向: 图像情感计算、模式识别; 刘佳欣(1996-), 女, 硕士研究生, 主要研究方向: 图像情感识别; 潘建达(1997-), 男, 硕士研究生, 主要研究方向: 图像情感识别。

收稿日期: 2021-10-04

法、纹理特征及抽象主体特写为主要特征的艺术图像相关情感研究不足。艺术图像作为艺术作品的重要表现形式,以显著抽象的视觉特征和显著主体向欣赏者传递情感语义信息,具有情景交融的特点。盛家川等将人类认知与 CNN 结合,根据中国画显著性区域和笔道复杂度提取感兴趣区域,利用预训练的 GoogLeNet 模型提取中国画的深层特征,结合中国画情感表达手法知识,微调模型结构训练参数,提升中国画情感识别的准确率<sup>[6]</sup>;王征等利用 AlexNet 模型提取中国画的深层特征并与颜色特征融合,利用 SVM 分类器进行情感分类<sup>[7]</sup>;张浩等以云南少数民族绘画作品为研究对象,分析了图像色相、亮度、饱和度 and 对比度对神经网络模型情感分类的影响,利用 Twitter 数据集提升 VGG16 模型的学习能力,采用微调的迁移学习策略实现对民族绘画作品的情感分类<sup>[8]</sup>。

结合目前图像情感识别领域的发展及艺术图像情感识别的现状,本文提出一种基于融合特征和迁移学习的艺术图像情感识别模型 HCFNet。首先,在 HSV 和 YCrCb 颜色空间下,分别提取经 CLAHE 处理后的 H-S 二维特征和 CrCb 二维特征,并分别输入对应模型 HNet 和 CrCbNet,进一步提取艺术图像的颜色特征;其次,基于 FPN 模型与 ResNet101 模型部分参数相同,为防止数据集规模较小出现过拟合现象,将 ImageNet 中预训练好的 ResNet101 部分参数迁移至 FPN 中,并对 FPN 进行改进,以提取艺术图像的深层情感语义特征;最后,将上述 3 种特征利用像素间加法进行融合,输入 BN 层和 Relu 激活函数中,构成本文提出的 HCFNet 模型。同时考虑在特征融合之前对两种颜色特征给予不同权重,达到模型训练最优状态,提升艺术图像情感识别的准确性。

## 1 相关工作

### 1.1 迁移学习

目前,深度学习已经广泛应用在图像情感识别领域中,有效解决图像底层特征和深层特征之间存在的“语义鸿沟”问题,但深度学习同时也需要海量数据集支撑模型的训练,而目前带有情感标签的艺术图像数据集较少,训练卷积神经网络模型时易出现过拟合现象。迁移学习的基本思想是将源领域(大型数据集)中训练好的模型权重转移至目标领域(小样本实验数据集)中结构相同的模型,并利用目标领域微调模型结构,该方法可有效防止深度学习在小样本数据集中出现过拟合现象,同时也

可以提升模型训练速度,缩短模型训练时间,减少模型参数<sup>[9]</sup>。在目前计算机视觉领域中,迁移学习已被广泛应用在图像分类、物体识别、目标检测、人脸识别等研究中,随着图像情感语义特征逐渐被关注,越来越多学者证明了将图像对象分类任务迁移至图像情感分类任务中的可行性。

迁移学习方法主要有两种形式,即特征提取和微调。特征提取是指利用预训练模型中全连接层之前的所有参数,提取目标领域中图像特征,并添加新的全连接层或其他分类器,以适应目标任务。而微调方法可以弥补小数据集造成训练不足产生过拟合的现象,通过固定部分卷积层参数提取图像普适特征,微调其它层参数用于训练目标数据集,避免重新训练模型造成参数多训练时间过长的缺陷。Yosinski 等冻结在 ImageNet 数据集中预训练的 CNN 模型部分网络层并进行迁移,该数据集约有 120 万张图像,1 000 种图像类别,主要分为物体和动物两大类,最后得出 CNN 模型提取的图像底层视觉特征具有普适性,而深层语义特征适用于与源领域类型相同或相近的小样本目标数据集中<sup>[10]</sup>。源领域和目标领域样本类别相似度越高,迁移学习效果越好;反之,源领域和目标领域差别越大,迁移学习效果将负方向发展。因此,在使用迁移学习方法进行图像情感分类识别时,可根据实际情况迁移全部或部分预训练模型的参数,并利用实验数据集对模型进行微调,以提升模型分类准确率,防止出现过拟合现象。

### 1.2 CLAHE 处理

颜色特征是图像最基本的底层视觉特征,蕴含最关键与最敏感的视觉信息。在对图像底层颜色特征研究过程中,多数学者习惯采用颜色直方图表达颜色特征,可直观看出图像像素分布情况。在同一数据集中,图像像素分布参差不齐,针对像素分布不均匀的图像,部分区域较整体图像相比较为明亮或光线较暗,因此采用普通颜色直方图均衡进行全局图像增强时效果不太理想。自适应直方图均衡(AHE)与颜色直方图相比,优势在于该方法通过计算图像每一处显著区域的直方图,重新调节图像亮度值进行均匀分布,以改善图像每个区域的局部对比度,增强图像边缘的清晰度,但 AHE 方法容易导致图像产生噪声,进而造成过度增强现象。而限制对比度自适应直方图均衡(CLAHE)与 AHE 相比,通过对每一像素邻域的对比如限制,得到对应的变换函数,可有效抑制图像噪声的增强<sup>[11]</sup>。其原理为在计算邻域的 CDF 前,通过定义阈值对直方图进行

裁剪,限制 CDF 及变换函数的斜率,直方图切割部分使用的定义阈值称作修剪限幅 (clip limit),产生的新直方图若仍然超过设定阈值,可适当增加阈值,重新调整,如图 1 所示。与原图像相比,经 CLAHE 方法处理后的图像细节处更为突出,特征更加丰富,颜色直方图较原图像相比也更为平滑,如图 2 所示。

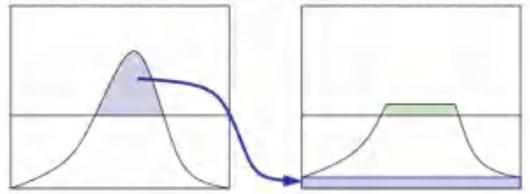


图 1 CLAHE 原理图

Fig. 1 The schematic chart of CLAHE

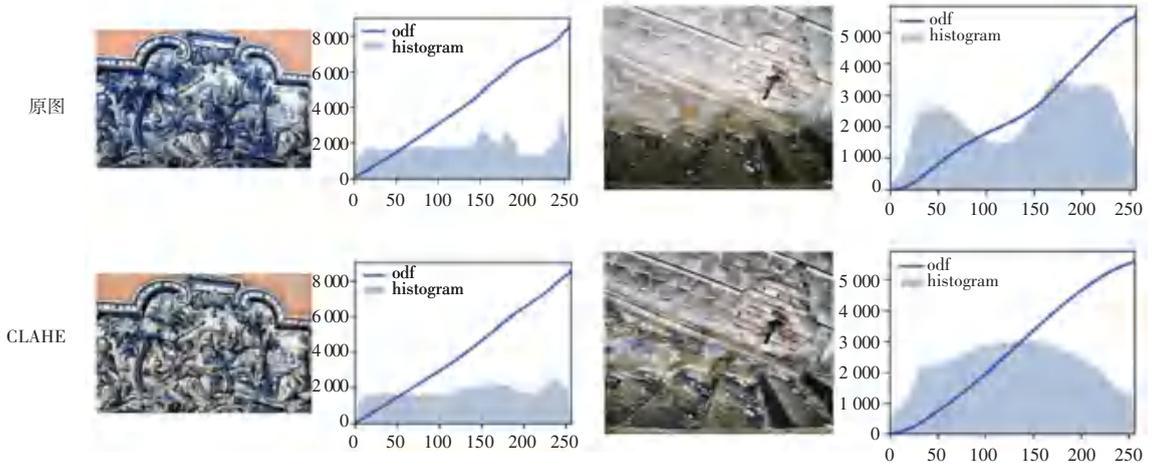


图 2 CLAHE 颜色特征可视化示例图

Fig. 2 The example graph of CLAHE color feature visualization

## 2 艺术图像特征提取

### 2.1 H-S 颜色特征

实验证明,在图像底层特征中,颜色特征更能体现图像的情感语义特征<sup>[2,8]</sup>。因此,本文实验将在 CLAHE 处理后提取图像颜色特征的基础上进行图像颜色特征提取。图像颜色空间除 RGB 空间外,还包括 HSV、YCrCb 空间等。大量心理学研究表明,HSV 颜色空间更贴近人眼视觉特征,其中  $H$  (Hue) 指色调,即日常描述的红、绿、蓝,用角度度量,取值范围为  $0 \sim 360$ ;  $S$  (Saturation) 指饱和度,即颜色接

近光谱度的程度,颜色越接近光谱色,则饱和度越高,进而表明颜色越纯,其取值范围为  $0.0 \sim 1.0$ ;  $V$  (Value) 指亮度,通常取值范围为  $0.0 \sim 1.0$ <sup>[12]</sup>。

因此,鉴于图像 HSV 空间具有更符合人类视觉对颜色感知的特点,首先将图像从 RGB 格式转为 HSV 格式,经 CLAHE 处理后利用色调-饱和度二维直方图(H-S)方法,提取图像中起显著作用的 H-S 颜色特征,该二维直方图中色调值 ( $H$ ) 介于  $0 \sim 180$  之间,饱和度 ( $S$ ) 介于  $0 \sim 256$  之间,输入图像对应的 H-S 二维直方图可视化效果如图 3 所示。

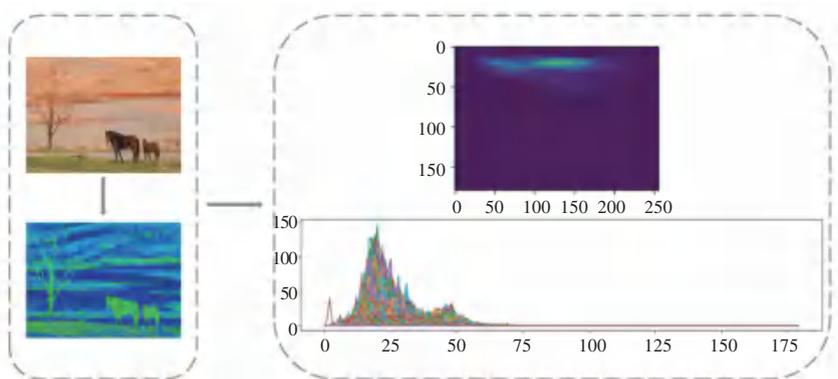


图 3 H-S 二维直方图可视化示例图

Fig. 3 The example graph of H-S two-dimensional histogram visualization

将 H-S 二维直方图提取到的颜色特征输入含有两个卷积层、两个最大池化层、一个全连接层、3 个 BN 层和 3 个激活层的网络中,其中卷积核大小为 3 中卷,  $filters = 64, stride = 2, padding = 1$ ;最大池化层大小均为 2 最大,  $stride = 2$ ;全连接层神经元个

数为 1 000;3 个激活层均使用 *Relu* 激活函数;卷积层和全连接层后均连接 BN 层,防止模型的过拟合。将该网络模型命名为 HSNet,输出结果即为图像的 H-S 颜色特征,记为  $X_1$ 。H-S 颜色特征提取过程如图 4 所示。

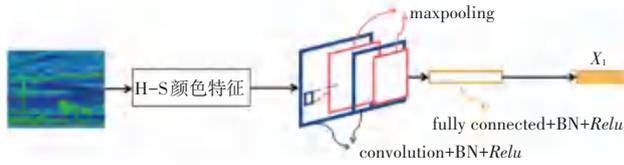


图 4 H-S 颜色特征提取

Fig. 4 H-S color feature extraction map

### 2.2 CrCb 颜色特征

图像颜色空间除常见的 HSV、RGB 外,还存在一种符合人类视觉特征的颜色空间,即 YCrCb,又称 YUV 颜色空间。YCrCb 主要用于优化彩色视频信号的传输,其中  $Y$  表示颜色的明亮度; $Cr$  表示色调,是颜色中红色部分的分量值; $Cb$  表示饱和度,是颜色中蓝色部分的分量值<sup>[13]</sup>。基于在 YCrCb 颜色空间中,对于图像的"色度"起决定性作用的是  $Cr、Cb$

两通道,因此该处在提取颜色特征时,只考虑 CrCb 二维空间特征。具体方法同上,将图像从 RGB 格式转为 YCrCb 格式,经 CLAHE 处理后利用色调-饱和度二维直方图方法,提取图像中起显著作用的 CrCb 颜色特征,在该二维空间特征中,色调值 ( $Cr$ ) 介于 16~224 之间,饱和度 ( $Cb$ ) 介于 16~224 之间。针对输入图像,对应 CrCb 二维直方图可视化效果如图 5 所示。

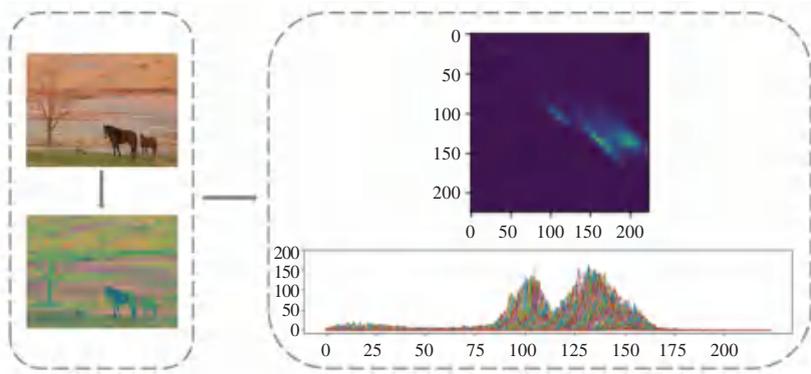


图 5 CrCb 二维直方图可视化示例图

Fig. 5 The example graph of CrCb two-dimensional histogram visualization

将提取到的 CrCb 二维直方图特征输入含有两个卷积层、两个最大池化层、一个全连接层、3 个 BN 层和 3 个激活层的网络中,其中卷积核大小为 3 中卷,  $filters = 64, stride = 2, padding = 1$ ;最大池化层均为 2 为大,  $stride = 2$ ;全连接层神经元个数为 1 000;

3 个激活层使用 *Relu* 函数;卷积层和全连接层后均连接 BN 层防止过拟合。将该网络模型命名为 CrCbNet,输出结果即为图像的 CrCb 颜色特征,记为  $X_2$ 。CrCb 颜色特征提取过程如图 6 所示。

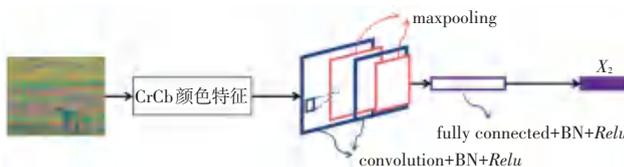


图 6 CrCb 颜色特征提取

Fig. 6 CrCb color feature extraction map

## 2.3 改进的 FPN 特征

2017 年, Lin 等提出了 FPN (feature pyramid networks) 特征金字塔结构, 该网络模型同时兼顾图像的底层特征和深层特征, 将低分辨率、高级语义信息的深层特征和高分辨率、低级语义信息的底层特征进行自上而下的侧边连接逐层融合, 构成图像多尺度的融合特征, FPN 结构如图 7 所示<sup>[14]</sup>。FPN 模型结构大致分为自底而上、自顶而下和横向连接 (lateral connection) 3 部分。在自底向上阶段, 采用卷积神经网络中 ResNet 模型作为网络主干结构, 将 conv2\_x、conv3\_x、conv4\_x、conv5\_x 各部分最后一个残差结构输出记为  $\{C2, C3, C4, C5\}$ ; 而在自顶向下阶段, 采用上采样内插值方法, 在原有图像像素点之间选取合适的插值算法, 插入新的元素, 进而扩大原图像的大小, 通过对特征图上采样, 使特征图具有和下一层的特征图相同的大小, 充分利用图像各层的细节信息。

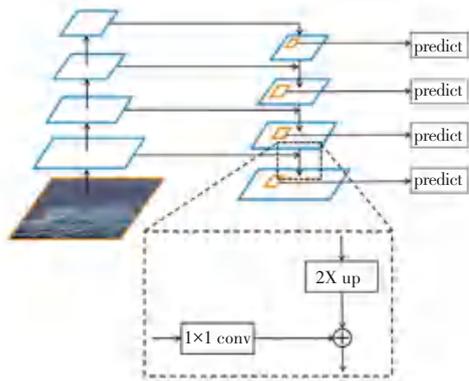


图 7 FPN 特征金字塔结构示意图

Fig. 7 FPN (feature pyramid networks) structure diagram

如图 8 所示, 自底向上的特征图  $C5$  经过 11 卷积核卷积得到  $P5$ , 之后通过上采样得到与  $C4$  大小一样的特征图, 再利用像素间加法将  $C4$  通过  $1 \times$  过卷积核卷积后的特征映射与  $P5$  上采样特征图相加。为了消除上采样产生的混叠效应, 利用  $3$  采样卷积核卷积得到  $P4$ , 重复迭代上述过程, 最终得到多尺度特征映射  $\{P2, P3, P4, P5\}$ 。

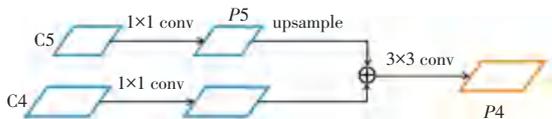


图 8 图像特征融合示意图

Fig. 8 Schematic diagram of Image features fusion

FPN 模型是在 ResNet 模型的基础上构建的, 两个模型部分结构参数相同, 为防止因数据集较小而出现过拟合的现象, 本实验将迁移学习思想与 FPN 模型结合, 截取部分 ResNet101 模型的预训练参数, 并将其应用于训练后续改进的 FPN 模型。

为充分利用 FPN 特征金字塔模型各层特征信息, 对 FPN 模型进行改进, 如图 9 所示。将  $\{P2, P3, P4, P5\}$  逐一输入到大小为  $1 \times 1$  的自适应的平均池化层中, 利用像素间加法原则将输出特征进行融合, 得到由改进 FPN 特征金字塔模型提取的整体特征  $P$ , 并将该特征先后输入全连接层、BN 层、激活层中, 输出结果记为特征  $X_3$ 。

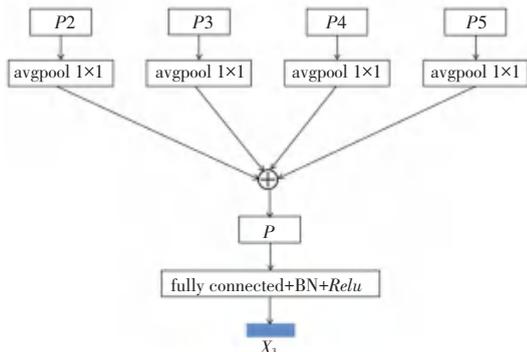


图 9 改进 FPN 特征提取模型

Fig. 9 Improved FPN feature extraction map

图像输入改进 FPN 模型后, 模型不同层所提取到的特征可视化效果如图 10 所示。随着模型层数的加深, 特征图效果越模糊, 但从模型第十二层开始, 随着层数加深, 提取到的图像特征开始逐层融合, 特征图变得更为直观具体, 情感语义特征更加丰富。

## 2.4 图像特征融合

将上述提取到的 H-S 颜色特征  $X_1$ , CrCb 颜色特征  $X_2$  及 FPN 多尺度融合特征  $X_3$  利用像素间加法进行融合, 相比于 concat 融合方式, 该融合方法计算时间短并且参数较少, 并将融合后的特征输入 BN 层和 ReLU 激活函数中。

## 3 算法主要流程

将本文实验的整体网络模型命名为 HCFNet, 其网络模型结构如图 11 所示。输入的图像大小为  $224 \times 224$ , 由于艺术图像的颜色特征处理需在 OpenCV 库中进行, 因此在提取 H-S 颜色特征和 CrCb 颜色特征之前需将 *tensor* 向量转换为 numpy 数组形式, 再各自进行颜色特征处理; 将提取到的二维 H-S 颜色特征和 CrCb 颜色特征转回张量形式, 并分别输入 HSNNet 和 CrCbNet 网络中, 进一步提取图像颜色特征, HSNNet 输出的特征记为  $X_1$ , CrCbNet 输出的特征记为  $X_2$ , 图像输入到改进 FPN 网络模型中得到深层语义特征记为  $X_3$ ; 最后, 将得到的图像底层特征和深层情感语义特征融合, 输入到 BN 层和 ReLU 激活函数中, 对艺术图像进行情感语义识别。

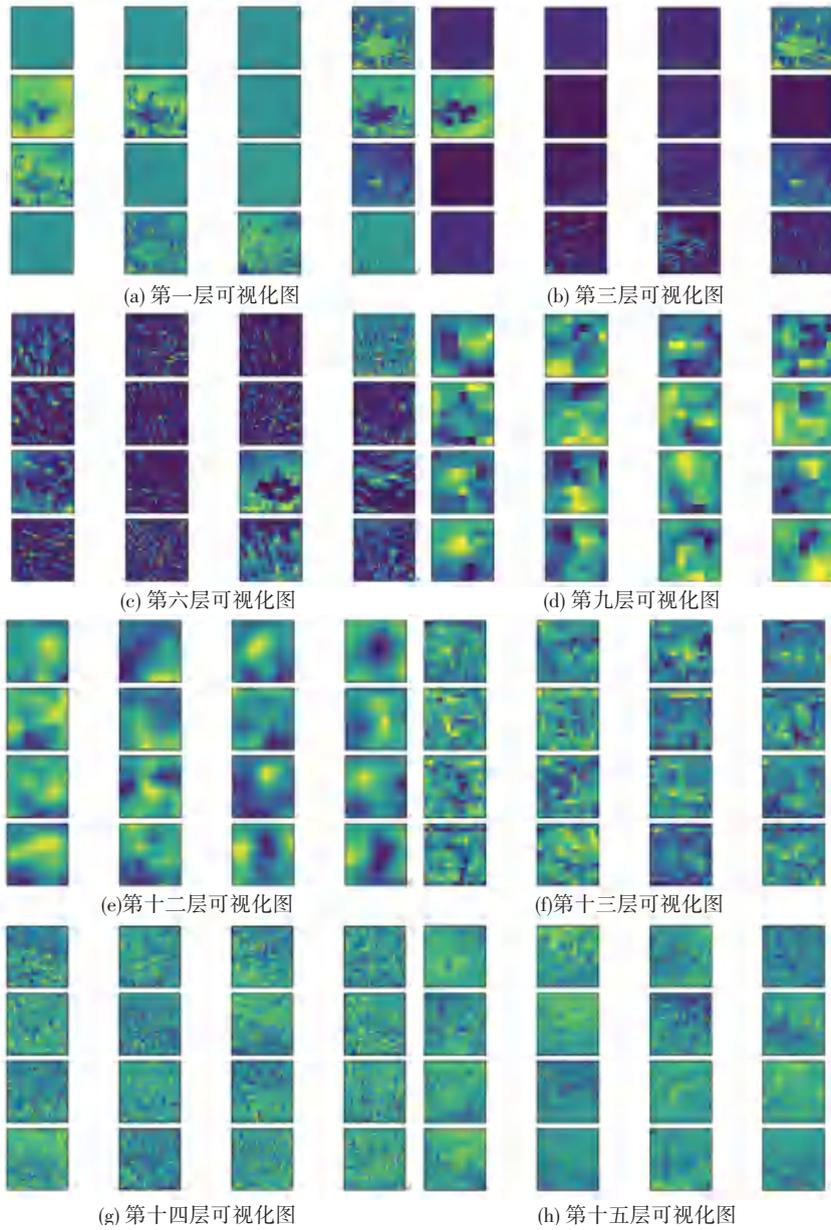


图 10 不同层深层特征可视化图

Fig. 10 Visualization of deep features in different layers

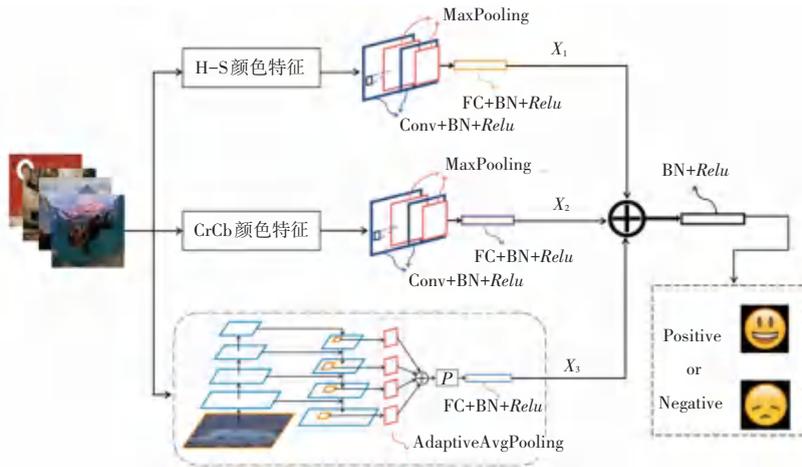


图 11 HCFNet 网络模型结构图

Fig. 11 HCFNet network model structure diagram

## 4 实验结果与分析

### 4.1 实验环境

本文实验环境:处理器为 Intel i5-8279U,主频 2.40 GHZ,实验基于 Python 3.7 的深度学习框架 Pytorch 运行,开发工具为 PyCharm,使用 Linux 操作系统,GPU 进行训练,GPU 为 NVIDIA Tesla V100-SXM2-16GB,显存总量为 16 160 MiB。

### 4.2 实验数据集

#### 4.2.1 数据集介绍

本文实验使用 Flickr 公开数据集,图像情感标签共分 3 类,分别为 Positive、Negative 和 Neutral。基于 Flickr 数据集中 Neutral 标签下图像数量较少,因

此只保留 Positive 和 Negative 两类情感特征,并人工筛选 Positive 和 Negative 标签下的图像各 1 920 张,作为本文实验数据集。人工筛选原则为选取具有显著视觉特征、真实情感标签符合原本标定的艺术类情感图像,以保证各类标签下样本图像数量相等,确保情感识别结果均衡。本文数据集各标签的示例图像如图 12 所示。

#### 4.2.2 数据集预处理

由于原数据集大多数图像格式是通道数为 3 的 RGB 图像,其中也存在少数灰度图像,其通道数为 1,为方便后续实验的展开,在数据预处理阶段统一将图像转为 RGB 图像。



(a) 积极情感的图像集



(b) 消极情感的图像集

图 12 Flickr 数据集示例图像

Fig. 12 The schematic images of Flickr Dataset

### 4.3 实验评价指标

实验评价指标为图像情感分类准确率 (Accuracy, Acc),表示如式(1):

$$Acc = \frac{N}{Total} \times 100\% \quad (1)$$

其中, Total 为测试集样本总数, N 为测试集中标签预测正确的样本数。

实验采用 K 折交叉验证方法,因此准确率最终为经 K 折交叉验证后准确率的平均值。

### 4.4 实验结果分析

为证明本文实验方法对艺术图像的情感识别任务有效,在训练过程中,采用五折交叉验证方法,默认迭代轮次为 20 轮,学习率设为 0.005,批大小设为 64,使用交叉熵损失函数,选择 SGD 优化器,动量梯度下降参数设置为 0.9。

在本文实验中,选择经典卷积神经网络模型 VGG16 的迁移学习模型、VGG16\_BN 的迁移学习模型以及未融合迁移学习 ResNet101 的改进 FPN 和本

文提出的模型算法进行比较,五折交叉验证后测试集的平均准确率见表 1。在各模型的五折交叉验证中,选取训练集预测结果最好的一折,并绘制该折训练全过程的准确率及损失值变化情况,各模型的准确率随着迭代次数变化的折线对比图如图 13 所示,各模型随迭代次数变化的损失变化折线对比情况如图 14 所示。

表 1 与经典卷积神经网络模型准确率对比

Tab. 1 Accuracy comparison with classic CNN models

模型名称	准确率 Acc/ %
VGG16(TL)	87.73
VGG16_BN(TL)	87.16
FPN(F)	77.76
FPN+ResNet101(TL)( $X_3$ )	90.16
$X_1 + X_3$	90.10
$X_2 + X_3$	89.77
$X_1 + X_2 + X_3$ (无权重 HCFNet)	<b>90.31</b>

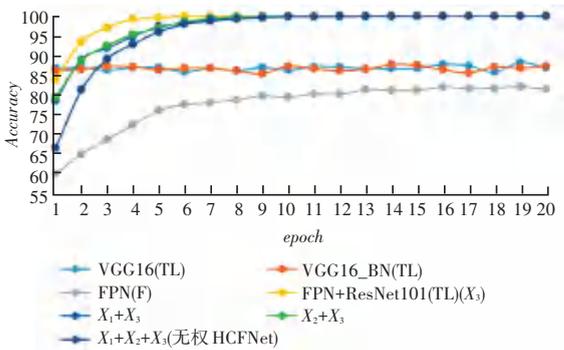


图 13 各模型的准确率变化折线图

Fig. 13 Line chart of accuracy changes of each model

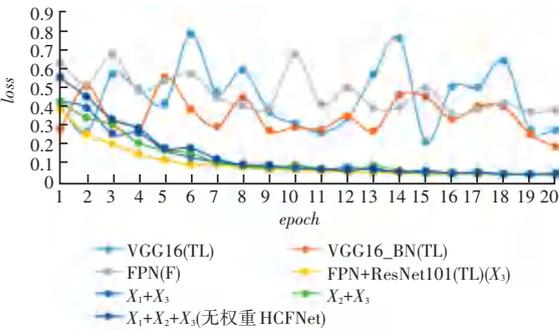


图 14 各模型的损失值变化折线图

Fig. 14 Line chart of loss value changes of each model

由表 1 可知,以改进的 FPN 模型特征  $X_3$  作为主干特征,分别用  $X_1 + X_3$ 、 $X_2 + X_3$  和  $X_1 + X_2 + X_3$  3 组特征组合,逐一对数据集进行训练,所得到的准确率较 VGG16 的迁移学习模型、VGG16\_BN 的迁移学习模型以及未融合迁移学习 ResNet101 的改进 FPN 模型算法均有所提升。将 H-S 和 CrCb 两个二

维颜色特征  $X_2$  和  $X_3$ ,分别与  $X_3$  特征融合,所得准确率有所降低;但是横向比较,H-S 二维颜色特征较 CrCb 二维颜色特征相比更能反应图像的情感语义信息,由此可说明 HSV 颜色空间更能直接影响图像情感特征。而本文所提出的将  $X_1$ 、 $X_2$  和  $X_3$  3 个特征融合的 HCFNet 模型准确率高达 90.31%,比未融合迁移学习 ResNet101 的改进 FPN 模型提升 12.55%,比本文改进的 FPN 模型特征  $X_3$  提升 0.15%。

结合图 13 和图 14 各模型在训练集的表现情况上看,在训练过程中,基于迁移学习的 VGG16 和 VGG16\_BN 模型准确率均保持在 85%左右,未融合迁移学习 ResNet101 的改进 FPN 模型,其准确率从 60%上升至 80%左右;而  $X_3$ 、 $X_1 + X_3$ 、 $X_2 + X_3$  和  $X_1 + X_2 + X_3$  (HCFNet) 各模型在  $epoch = 8$  时,准确率已经达到 100%。在损失值变化情况方面,基于迁移学习的 VGG16、VGG16\_BN 和未融合迁移学习 ResNet101 的改进 FPN 模型的  $loss$  值损失波动幅度较大;相反,  $X_3$ 、 $X_1 + X_3$ 、 $X_2 + X_3$  和  $X_1 + X_2 + X_3$  (HCFNet) 模型的收敛速度非常快,当  $epoch = 13$  时,损失值趋近平缓。

基于 H-S 特征  $X_1$  和 CrCb 特征  $X_2$  均属于颜色特征范畴,在特征融合前尝试对  $X_1$  和  $X_2$  给予权重,将给予权重的两个特征与  $X_3$  再进行融合,形成加权融合特征  $FX$ , 式(2):

$$FX = [a * X_1, b * X_2, X_3] \quad (2)$$

其中,权重系数和为 1。

给予不同权重后 HCFNet 模型的测试结果见表 2。由表 2 可知,对  $X_1$  特征和  $X_2$  特征给予权重后,随着  $X_1$  特征权重系数的增加, HCFNet 模型的准确率逐渐提升,当权重系数  $a = 0.9, b = 0.1$  时, HCFNet 模型准确率最高,能达到 90.63%,较未加权的 HCFNet 模型结果提升 0.32%。

表 2 加权 HCFNet 准确率对比

Tab. 2 Weighted HCFNet accuracy comparison

模型名称	准确率 Acc/ %
$X_1 + X_2 + X_3$ (无权重 HCFNet)	<b>90.31</b>
$0.6 * X_1 + 0.4 * X_2 + X_3$	89.58
$0.7 * X_1 + 0.3 * X_2 + X_3$	90.29
$0.8 * X_1 + 0.2 * X_2 + X_3$	90.31
$0.9 * X_1 + 0.1 * X_2 + X_3$	<b>90.63</b>

综上所述,本文提出的基于融合特征和迁移学习的艺术图像情感识别方法能有效地识别艺术图像情感语义特征,在改进 FPN 模型中引入迁移学习思想,提取艺术图像的多尺度深层语义特征,在此基础

上融合图像的颜色特征,并对不同颜色特征加权,可更好地弥合艺术图像底层特征和深层特征之间的“语义鸿沟”。

## 5 结束语

本文提出一种基于融合特征和迁移学习的艺术图像情感识别模型 HCFNet,充分利用艺术图像的不同颜色空间下的二维颜色特征,即提取 H-S 颜色特征和 CrCb 颜色特征;将迁移学习思想引入 FPN 特征金字塔网络中并进行改进,以多角度提取图像深层语义特征;将 3 种特征融合构成本文模型 HCFNet。考虑在特征融合之前对两种颜色特征加权,该模型情感识别准确率可达到 90.63%,证明本文方法可有效实现艺术图像情感识别任务。

## 参考文献

- [1] KHANNA, PRITEE, YOKOTA, et al. Clustering of hierarchical image database to reduce inter- and intra-semantic gaps in visual space for finding specific image semantics [J]. *Journal of visual communication & image representation*, 2016, 38 (7): 704-720.
- [2] 李志义,许洪凯,段斌. 基于深度学习 CNN 模型的图像情感特征抽取研究[J]. *图书情报工作*, 2019, 63(11): 96-107.
- [3] 杨子文,陈蕾,浦建宇. 基于两层迁移卷积神经网络的抽象图像情感识别[J]. *中国科学技术大学学报*, 2019, 49(1): 40-48.
- [4] 蓝亦伦,孟敏,武继刚. 基于视觉语义联合嵌入和注意力机制的情感预测[J]. *计算机科学*, 2020, 47(11): 250-254.
- [5] 杨文武,普园媛,赵征鹏,等. 基于多层特征描述及关系学习的智能图像情感识别[J]. *陕西师范大学学报(自然科学版)*, 2019, 47(5): 40-48.
- [6] 盛家川,陈雅琦,王君,等. 融合人类认知网络优化的中国画情感识别[J]. *模式识别与人工智能*, 2020, 33(2): 141-149.
- [7] 王征,李皓月,许洪山,等. 基于卷积神经网络和 SVM 的中国画情感分类[J]. *南京师大学报(自然科学版)*, 2017, 40(3): 74-79, 86.
- [8] 张浩,徐丹. 基于深度学习的少数民族绘画情感分析方法[J]. *中国科学:信息科学*, 2019, 49(2): 204-215.
- [9] HAN D, LIU Q, FAN W. A new image classification method using CNN transfer learning and web data augmentation[J]. *Expert Systems with Applications*, 2018, 95(2): 43-56.
- [10] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks[C]//27<sup>th</sup> International Conference on Neural Information Processing Systems, 2014: 3320-3328.
- [11] REZA A M. Realization of the contrast limited adaptive histogram equalization (CLAHE) for real-time image enhancement [J]. *Journal of VLSI signal processing systems for signal, image and video technology*, 2004, 38(1): 35-44.
- [12] 刘喜姣. 敦煌壁画中祥云文化元素的特征提取及演化规律研究[D]. 武汉:武汉理工大学, 2014.
- [13] 曹建芳. 场景图像情感语义分析关键技术研究[D]. 太原:太原理工大学, 2015.
- [14] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition, Honolulu: IEEE, 2017: 2117-2125.
- [15] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [J]. *arXiv preprint arXiv: 14091556*, 2014.
- [16] ŞAH M, DIREKOĞLU C. Evaluation of image representations for player detection in field sports using convolutional neural networks [C]//proceedings of the International Conference on Theory and Applications of Fuzzy Systems and Soft Computing, F, 2018.
- [17] LU K, CHEN J, LITTLE J J, et al. Light cascaded convolutional neural networks for accurate player detection [J]. *arXiv preprint arXiv:170910230*, 2017.
- [18] 张丽晶. 足球比赛视频中球员检测方法的研究 [D]. 北京:北京理工大学, 2018.
- [19] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single Shot MultiBox Detector [C]//European Conference on Computer Vision Springer, Cham, 2016: 21-37.
- [20] 吕涛. 足球视频中球的检测与跟踪算法研究 [D]. 西安:西安电子科技大学, 2012.
- [21] 吕泉. 足球视频目标检测与跟踪方法研究 [D]. 哈尔滨:哈尔滨工业大学, 2018.
- [22] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need [C]//proceedings of the Advances in neural information processing systems, 2017.
- [23] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [24] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]// Proceedings of the European Conference on Computer Vision, 2020.
- [25] JIANG Y, CUI K, CHEN L, et al. SoccerDB: A Large-Scale Database for Comprehensive Video Understanding [C]// Proceedings of the 3<sup>rd</sup> International Workshop on Multimedia Content Analysis in Sports. 2020: 1-8.
- [26] LIN T Y, MAIRE M, BELONGIE S, et al. Microsoft COCO: Common Objects in Context [J]. Springer International Publishing, 2014.
- [27] LOSHCHELOV I, HUTTER F. Fixing weight decay regularization in adam [J]. 2018.
- [28] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]// proceedings of the Proceedings of the IEEE conference on computer vision and pattern recognition, 2016: 770-778.

(上接第 145 页)