

基于深度学习的自适应水墨画计算美学评估

张佳婧¹⁾, 于金辉²⁾, 缪永伟¹⁾, 彭韧³⁾

¹⁾ (浙江理工大学信息学院 杭州 310018)

²⁾ (浙江大学 CAD&CG 国家重点实验室 杭州 310058)

³⁾ (浙江大学计算机科学与技术学院 杭州 310027)

(zhangjj@zstu.edu.cn)

摘要: 近几年艺术作品的计算美学评估已成为一个热门的研究方向。但现有工作主要研究照片和油画, 关于水墨画的定量审美评估却鲜有尝试。水墨画通过水墨相调和笔法变化来表现画面, 因而在视觉特征、语义特征和审美准则上与照片和油画有显著不同。针对此问题, 采用深度学习技术, 提出一种自适应的水墨画计算美学评估框架。该框架首先构建水墨画图像美学评价基准数据集; 然后根据水墨画审美标准提取全局与局部图像块作为多路输入, 并设计一种多视角并行深度卷积神经网络来提取深度审美特征; 最后基于水墨画的题材查询机制, 构建自适应深度审美评估模型。实验结果表明, 文中包含6个并行题材卷积组的多视角网络架构相较基础VGG16架构有较高的审美评估性能, 提取的深度审美特征明显优于传统手工设计特征, 其自适应模型评估结果与人工审美评价之间达到0.823的皮尔森高度显著相关, 且均方误差为0.161。此外, 干扰实验表明, 文中的网络对构图、墨色和纹理3个绘画要素较为敏感。该研究将不仅为国画计算美学评估提供了一个基于深度学习的参考框架, 而且有助于进一步探索人类审美感知与水墨画中深度学习特征之间的关系。

关键词: 深度学习; 水墨画; 计算美学评估; 多视角并行深度神经网络; 自适应模型

中图分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2021.18815

Self-Adaptive Computational Aesthetic Evaluation of Chinese Ink Paintings Based on Deep Learning

Zhang Jiajing¹⁾, Yu Jinhui²⁾, Miao Yongwei¹⁾, and Peng Ren³⁾

¹⁾ (Department of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018)

²⁾ (State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

³⁾ (College of Computer Science and Technology, Zhejiang University, Hangzhou 310027)

Abstract: Computational aesthetic evaluation of artworks has become an active research direction in recent years. However, current works mainly focus on oil paintings and photographs, there have been few attempts in quantitative aesthetic evaluation of Chinese ink paintings. Chinese ink painting uses ink blended with water and a variation of brushwork to depict picture, which differs significantly from photographs and oil paintings in visual features, semantic features, and aesthetic principles. Aiming at this problem, we propose a framework of self-adaptive computational aesthetic evaluation of Chinese ink paintings based on deep learning technique. Firstly, we build an aesthetic evaluation standard dataset for ink painting images. Secondly, according to aesthetic principles of Chinese ink paintings, we design a multi-view parallel deep neural

收稿日期: 2020-11-27; 修回日期: 2021-01-27. 基金项目: 浙江省自然科学基金(LQ20F020022); 国家自然科学基金(61772463, 61972458); 浙江理工大学科研启动基金(18032115-Y). 张佳婧(1991—), 女, 博士, 讲师, 主要研究方向为计算美学、多媒体计算、深度学习; 于金辉(1960—), 男, 博士, 教授, 博士生导师, 主要研究方向为计算美学、计算机动画、数字艺术; 缪永伟(1971—), 男, 博士, 教授, 博士生导师, CCF 会员, 主要研究方向为计算机视觉、数字图像处理; 彭韧(1967—), 男, 博士, 副教授, 硕士生导师, 主要研究方向为工业设计、媒体与交互设计.

network by taking global images and local patches as multi-column inputs to extract deep aesthetic features. Finally, we build a self-adaptive deep aesthetic model of Chinese ink paintings based on subject query mechanism. Experimental results show that, compared with the basic VGG16 architecture, our multi-view network that contains six paralleled subject convolution groups has higher aesthetic evaluation performance, the deep aesthetic features outperform the traditional hand-crafted features, and our proposed self-adaptive model can predict human aesthetic decision with highly significant Pearson correlation of 0.823, with mean squared error of 0.161. Moreover, interference experiments show that our network is sensitive to painting factors including composition, ink color, and texture. Our work not only offers a deeply-learned-based reference framework for quantitative aesthetic evaluation of Chinese paintings, but also reveals the relationship between human aesthetic perceptions and deeply-learned features extracted from Chinese ink paintings.

Key words: deep learning; Chinese ink painting; computational aesthetic evaluation; multi-view parallel deep neural network; self-adaptive model

水墨画是中国传统绘画艺术中的典型代表,通过水墨相调和笔法变化营造墨色干湿浓淡的层次变化,有如兼五彩的艺术效果和极高的美学价值.传统的水墨画审美评估只能通过鉴赏专家定性描述,其结果具有较大的主观不确定性.此外,互联网中存储的水墨画图像数量巨大,邀请为数不多的高水平专家以人工方式对它们逐一评估也不现实.如能用计算机高效、自动地对它们进行定量评估,将有助于帮助初学者在水墨画学习过程中对作品进行美感评价,对水墨画教学有重大指导意义,同时对水墨画高级检索和数字国画艺术馆的推荐展示等领域也有广阔应用前景.

计算美学是最近几年才出现的一个研究方向,即使计算机能模拟人类视觉系统和审美思维,并在相关应用中自动做出可行审美决策的计算方法研究^[1-2].目前该工作大部分以照片和西方绘画作为研究对象,以图像美感评价数据集为基础^[3-6],或人工设计视觉特征^[7-13],或近几年使用深度卷积神经网络自动提取高层美学特征^[14-23],而后者在审美分类与评价任务中取得了优越性能,但它们不能完全适用于水墨画的美感定量评估中,挑战有如下 3 点.

(1) 现有的美感评价数据集大都基于照片或西方绘画图像构建,而没有公共可用的针对水墨画美感评估的人工标定方案和图像数据.需要根据水墨画审美准则和自身特征从专业角度建立新的图像美学评价数据集.

(2) 水墨画在审美准则上讲究气韵生动、笔墨神韵、骨法用笔、随类赋彩和经营位置等^[23-26],这些评价标准对应高层语义,抽象程度高,难以用手

工特征来表示,易造成语义鸿沟.此外,水墨画在颜色、纹理和构图等视觉特征上与常见艺术形式有显著不同,现有方法难以准确和全面地描述水墨画审美属性.

(3) 不同水墨画题材对应不同的审美评估标准,例如,花鸟画通过色彩美抒发情感;山水画则强调笔墨技法,擅用散点透视法造势.此外,同一题材的水墨画根据内容其审美属性也相应变化,单一审美模型难以适用于所有题材类型.现有方法大多依据特定类型场景或先验知识训练独立的审美模型,具有一定的场景局限性.

为了解决上述问题,本文以传统水墨画作为研究对象,结合美术专业鉴赏理论,提出一种基于深度学习的自适应美感定量评估方法.首先收集不同题材类型的近现代国画艺术家专业作品及业余学生作品组成水墨画图像数据,并基于艺术专业人员评分构建水墨画美学评估基准数据集,进而为美学特征深度学习和模型构建提供了训练和测试数据.然后根据水墨画审美特点提取水墨画全局图像与自适应图像块作为多路输入,设计一种多视角并行深度卷积神经网络来提取深度审美特征,从而能更好地捕捉和量化水墨画中不同题材墨色、笔触和构图的审美感知信息.此外,本文将 AlexNet 在 ImageNet 上预训练的模型,针对水墨画题材语义进行迁移学习,进而提取微调后的网络输出作为视觉表示特征,并基于题材内容查询机制,构建自适应的水墨画深度审美评估模型.

该工作是将深度卷积神经网络应用于水墨画计算美学评估的首次尝试,为国画定量审美评估提供了一个基于深度学习技术的参考框架,是中

国传统艺术在当下人工智能时代的创新探索,具有重要研究意义和科学价值。

1 相关工作

图像计算美学评估是人工智能和计算机视觉领域中近几年的一个研究热点。本文主要针对图像美学评价基准数据集、图像美感定量评估相关工作进行介绍与分析。

1.1 图像美学评价数据集

在人工主观评价图像美学方面, Luo 等^[3]构建了含 17 690 张照片的 CUHK-PQ 数据集,带有二元审美标签,并被分组成 7 个场景类别,即“动物”“植物”“静物”“建筑”“景观”“人物”和“夜景”。Murray 等^[4]构建了一个面向图像视觉美学质量分析与度量的大型图像数据库(aesthetic visual analysis, AVA),采用在线图像共享评分方法,数据集包含 255 530 幅图像,每幅图像由人类艺术家会员评分,包含 60 多种摄影风格和场景语义标记。Sartori 等^[5]收集 500 幅艺术博物馆收藏的专业抽象画和 500 幅在线社交网络上传的业余抽象画作为数据集,并邀请 100 位参观者对每幅作品的第一印象进行评分,其中 1 代表非常消极情感,而 7 代表非常积极情感。Amirshahi 等^[6]在谷歌艺术项目上收集了 1 600 幅高质量的西方彩色油画,并对这些作品进行艺术时期、风格和主题等信息标注,建立了 JenAesthetics 数据集。与上述工作不同,本文通过收集艺术专业人员标定数据,构建了针对水墨画的人工审美评价数据集。

1.2 图像美感定量评估

早期的美学特征人工设计方法研究成果都是采用人工设计的低层特征来量化实现具体的摄影技巧或审美准则^[7]。王朝晖等^[8]通过提取视觉结构、暗通道和脸部区域特征对肖像照片进行美感评估。Dhar 等^[9]通过提取低层特征对构图、场景内容和自然室外天空光照 3 种高层属性进行评估,并使用这 3 种属性来预测照片美感质量和兴趣性。Obrador 等^[10]提出了一种基于内容的照片美感质量评估方法,根据不同照片内容将照片分为 7 类,包括风景、植物、动物、夜景、人像、静物和建筑,并根据不同类别的评估标准设计相应的主体区域和视觉特征提取算法。Su 等^[11]提出了一种基于“保持美感特征袋”的风景照片实时视图推荐系统。然而,水墨画没有照片里的简洁性、真实度、清晰度、视彩度、曝光度、景深、模糊效果以及相机基本摄影

技巧等因素影响,也没有肖像照片中的人脸区域特征,风景照片中专业的色调组合、室内外、天空光照等因素,所以上述方法并不完全适用于水墨画的美感定量评估。

在西方绘画方面, Li 等^[12]通过提取颜色分布、亮度、模糊效果和边缘分布等全局特征,以及分割区域的形状信息和颜色对比特征对油画审美视觉质量进行分类。Sartori 等^[5]通过提取 LAB 颜色和 SIFT 纹理描述子来区分积极与消极美感的抽象油画。但是,由于油画色彩丰富,强调光影的表现;抽象画主要通过形状和颜色以强烈的形式构成来表达,而水墨画色彩数量少,不拘于焦点透视,不强调自然界对于物体的光色变化,更多的是用墨色浓淡干湿变化和留白来表现层次,因此,上述方法均不能准确和全面地描述水墨画的审美属性。

在水墨画美感评估方面,张佳婧等^[13]提出一种基于低层特征的水墨画美感线性回归模型。但其只以齐白石花鸟画题材作为数据集,水墨画训练样本数量很少且缺乏题材和风格的多样性,未考虑不同题材内容对美感评估结果的影响,使得模型泛化评估能力较弱;此外其低层手工设计难以准确量化水墨画高层审美语义。

最近,随着深度学习技术的迅猛发展,一些美学特征深度学习方法研究工作通过改造用于图像识别的卷积神经网络(convolutional neural networks, CNN)^[14],来训练照片标定数据并自动学习有效的审美特征,其准确率与人工设计特征相比有较大提高。Lu 等^[15]提出了一种基于深度学习的照片审美打分模型(rating pictorial aesthetics using deep learning, RAPID),其中最后一层输出审美二元分类,并设计成双路视角以共同学习全局和局部图像特征。Tian 等^[16]提出一种依赖查询类别的深度审美模型,但其未考虑局部视角对照片整体美感评估的影响。Wang 等^[17]通过设计多场景卷积层训练深度学习模型来评估照片。Kao 等^[18]将场景信息融合到网络中,提出一种基于多任务关系学习框架的深度审美模型。Ma 等^[19]提出了一种自适应布局多图像块的深度卷积网络架构来评估照片美感。Talebi 等^[20]采用网络输出与真实直方图之间的距离作为损失函数进行训练,并提出一种 NIMA 模型来预测人类评估意见分布。Michal 等^[21]将人工设计的专业摄影特征与深度美学特征相融合进行审美预测和分类。李超等^[22]以素描作品作为研究对象,将深度卷积特征用于素描作品的分类与评价中。但是,上述网络结构大多数基于照片美学

数据集进行训练,其模型具有一定的场景局限性.此外,传统深度神经网络输入大都根据照片布局进行随机裁剪,其输入图像以及网络结构不能准确地量化水墨画的审美属性.本文将根据水墨画自身特点设计网络架构及网络输入,以达到同时保留全局和局部细节,并针对不同题材的水墨画设计自适应审美模型.

2 本文框架

图 1 显示了基于深度学习的自适应水墨画计算审美评估框架.首先收集 1 200 幅专业画家和业余学生的水墨画作品图像,并邀请艺术专业人员进行评分来构建美学评估基准数据集.然后根据水墨画审美特点改造基础 VGG16 网络结构,通过提取水墨画全局图像与自适应多图像块作为多路输入,设计一种多视角并行深度卷积神经网络来提取深度审美特征.最后对于每幅测试图像将 AlexNet 在 ImageNet 上预训练的模型针对水墨画题材语义进行迁移学习,进而提取微调后的网络输出作为视觉表示特征,并在整个训练集中根据相似度构建基于题材查询的子训练集,之后在该子训练集上,利用深度审美特征训练得到自适应的水墨画审美评估模型.

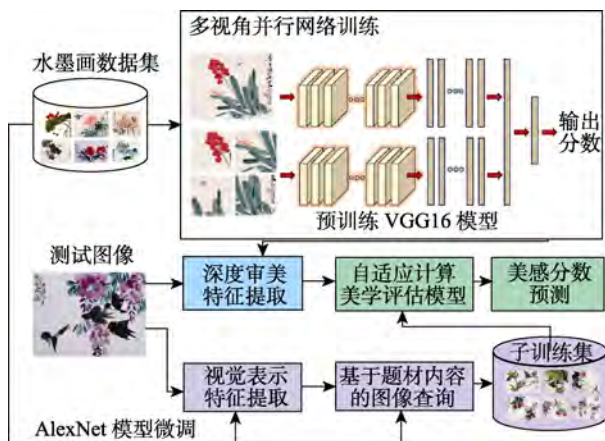


图 1 基于深度学习的自适应水墨画审美评估框架

作为一个基于数据驱动的问题,水墨画审美特征的深度学习需要依赖训练数据,因此数据集的构建就成为关键前提.下面将详细介绍水墨画美学评估基准数据集构建方法.

3 水墨画美学评估基准数据集

在水墨画图像数据收集上,本文拟选取 78 位

近现代著名国画艺术家,如齐白石、李可染、吴昌硕、王雪涛、潘天寿和张辛稼等的 600 幅专业水墨画作品,以及 600 幅业余画家如美术教学中的学生作品,其覆盖花鸟草虫、虾蟹、蔬果、走兽、山水和人物 6 类题材,将总共 1 200 幅水墨画图像作为数据集,部分示例如图 2 所示.其中一部分图像来自中国美术学院网上公开平台和“台北故宫在线数据库”,在类型和大小上进行认真选择,并以 JPEG 格式下载,收集了近 100 000 幅国画专业作品和数字文化遗产作品,可以公开访问和下载;另一部分图像通过扫描画册而得,如《中国现代水墨画》《水墨大观》《中国绘画名作欣赏》等知名画家作品集.之后对这些画作进行基本信息标注,包括画作名称、作者名称、创作年份、图像大小和题材类型.

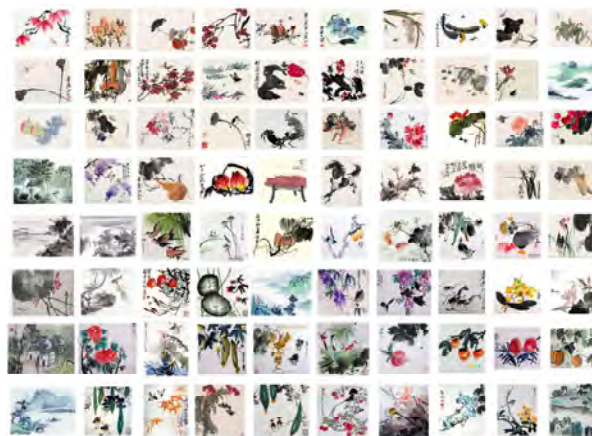


图 2 水墨画数据集示例

在人工评估测试阶段,本文邀请浙江理工大学艺术与设计学院、中国美术学院中国画与书法艺术专业的专家教授、普通教师、国画专业研究生以及艺术博物馆参观者(年龄范围是 18~65 岁),总共 180 名被试者参与评估打分.根据不同类别标记人员教育背景和艺术素养的不同,本文设置不同的权重并计算最终的加权平均分数,其中专家教授为 3,普通教师为 2,研究生和参观者为 1.

水墨画作为国画的一种重要表现形式,其欣赏和评价标准来自南朝谢赫在《古画品录》中归纳的“六法论”,这是千百年来被绘画界和书法界所公认的,也是中国古代品评水墨画的重要美学标准和原则^[26].基于上述审美标准,本文设计了一个评估打分页面.如图 3 所示,界面上方显示水墨画作品,每位测试者要求对每幅作品的以下几个方面进行评分:“整体美感”用来描述测试者对作品的整体印象,“墨色搭配”描述画面中色墨交融自然

度和浓淡层次变化,“笔触变化”描述画面中墨迹干湿对比度和行笔流畅性,“构图造势”描述画面中物体和留白的空间组织协调性. 其打分项采用最常见的 7 级量表进行打分, 范围为 1~7, 7 代表最优分数, 1 则代表最差分数. 所有图像被随机分成 6 组, 每组 200 幅, 每幅作品都会得到 30 个不同测试者评分. 当所有水墨画作品完成评分后测试结束, 每幅图像的加权平均值作为最终分数, 并用于深度神经网络模型训练.



图 3 人工评估打分页面

4 多视角水墨画深度审美特征提取

理解并量化水墨画的高层审美语义是弥补低层视觉特征与高层美学评估间鸿沟的重要环节, 而深度 CNN 在提取高层语义特征方面具有杰出的能力, 其高层输出可以视为高等的审美属性描述子. 因此, 本文根据水墨画审美特点改造基础 VGG16 网络结构, 设计了多视角并行深度 CNN 来提取审美特征.

4.1 网络基础架构

本文将水墨画审美评分任务表示为预测连续分数的回归建模问题. 将水墨画输入图像表示为 x_n , 并构造一个深度 CNN 来自动学习回归映射函数 $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, 从而预测美感分数 $\hat{y}_n = f(x_n)$. 给定水墨画训练数据 $\{(x_n, y_n)\}_{n \in [1, N]}$, 其中, N 为训练集大小, y_n 为人工基准评分, \mathbf{W} 为网络模型参数集合. 通过最小化欧几里得回归损失函数来最优化整个网络,

$$l(\mathbf{W}) = \frac{1}{2N} \sum_{n=1}^N \|\hat{y}_n - y_n\|_2^2 + \lambda \|\mathbf{W}\|_2^2 \quad (1)$$

其中, $\lambda \|\mathbf{W}\|_2^2$ 表示权重衰减正则化; λ 表示正则强度系数. 为了克服较小水墨画数据集的过拟合问题, 采用 VGG16 在 ImageNet 数据集上的预训练模型^[27]作为基准网络进行迁移学习^[28], 如图 4 所示. 该网络结构包括 13 个卷积层和 3 个全连接层, 固定前 12 个卷积层结构, 而对其他层进行如下调整:

Step1. 参考文献[17]中的多场景层设计了一种题材卷积层来替换原始网络第 13 层, 该题材层由 6 个并行的卷积网络组构成(每组 64 个卷积核), 分别用来提取水墨画中花鸟草虫、虾蟹、蔬果、走兽、山水、人物 6 类不同题材的审美描述符, 该层通过 6 个独立分支与第 12 层相连, 其输出通过平均池化与第 1 个全连接层相连, 从而使得网络针对不同的水墨画题材有较强的自适应学习能力.

Step2. 用包含 1 个神经元的回归层替换最后一个全连接层的 1 000 维分类概率来预测美感分数, 并使用欧几里得损失层替换掉最后输出的 softmax 损失层.

Step3. 将前 2 个全连接层的神经元数量从 4 096 减少到 512 和 256, 从而减少参数数量、防止过拟合, 并有利于特征级联. 经过修改后, 整个网络基础架构如表 1 所示.

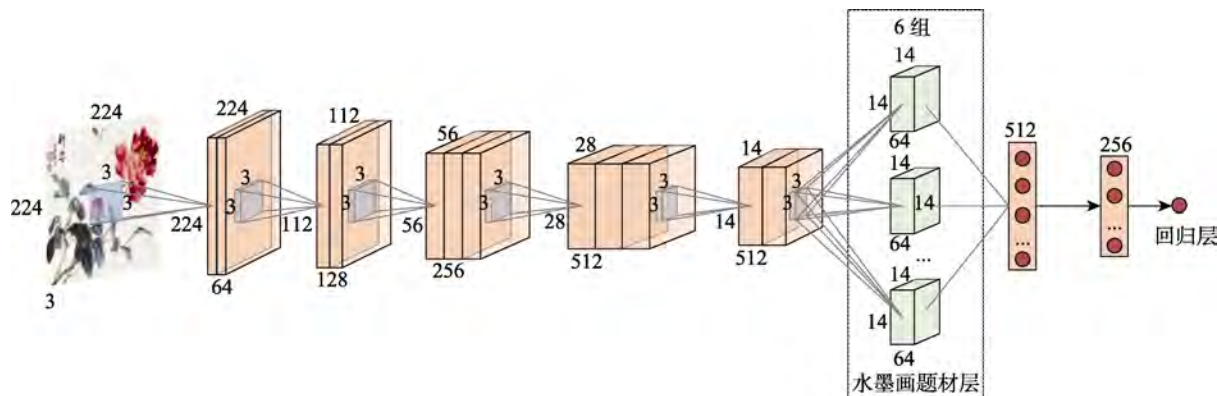


图 4 深度 CNN 基础架构

表 1 深度 CNN 基础架构

层数	核大小(步幅)	特征图大小
conv1-2	3×3(1)	224×224×64
max-pool1	2×2(2)	112×112×64
conv3-4	3×3(1)	112×112×128
max-pool2	2×2(2)	56×56×128
conv5-7	3×3(1)	56×56×256
max-pool3	2×2(2)	28×28×256
conv8-10	3×3(1)	28×28×512
max-pool4	2×2(2)	14×14×512
conv11-12	3×3(1)	14×14×512
conv13(6组)	3×3(1)	14×14×64
fc1		1×1×512
fc2		1×1×256
regression		1×1×1

为了能根据不同水墨画题材更有效地提取审美特征, 本文设计了一种预训练方法来初始化模型. 在网络参数初始化阶段, 固定 VGG16 模型前 12 个卷积层预训练好的权重, 作为网络中对应层

的初始化参数, 而针对第 13 个卷积层, 使用相应题材训练数据对每个卷积组进行单独微调, 其相应网络权重也不断更新. 当所有组都训练完后, 每组训练权重被并行链接到水墨画题材卷积层来进行初始化, 从而确保每个题材类别都有自己相应的神经元激活. 最后 2 个全连接层和回归层参数则采用随机初始化.

4.2 多视角并行深度 CNN

水墨画审美感知需要综合考虑整体和局部视角, 例如山水画在整体构图上讲究“五字法”和“三远法”, 而在局部上讲究留白分布和疏密对比.

通过扩展单路基础网络结构, 本文设计了一种多视角并行深度 CNN 架构, 如图 5 所示. 首先根据水墨画审美准则和基本结构提取不同表达形式的全局图像, 并自适应地选择图像块作为多路输入; 然后各路输入经过各自 VGG16 网络提取的特征, 并通过统计聚合结构合并; 最后将输出的全局和局部特征级联, 从而进行审美预测.

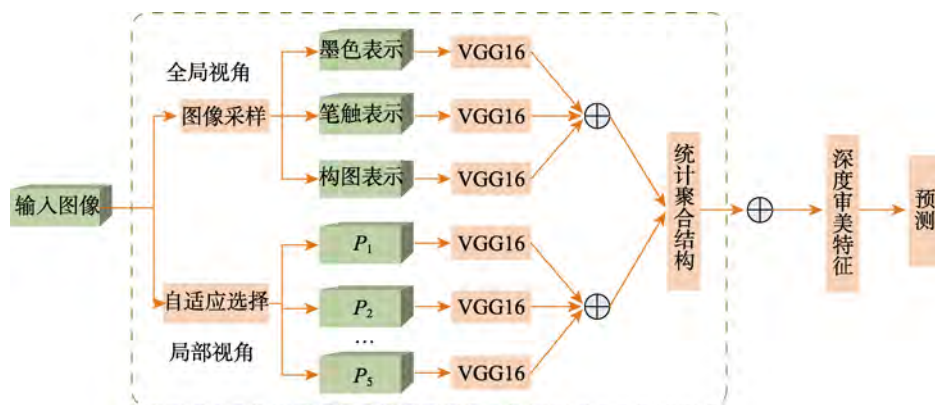


图 5 多视角并行 CNN

4.2.1 水墨画全局和局部视角输入

对于全局视角输入, 在墨色上, 水墨画通过彩色墨迹笔触、留白之间的对比, 以及浓淡灰度变化来表现. 这里选取水墨画原图、HSV 图的 H 和 S 通道、灰度图作为输入; 在笔触上, 因图像的小波系数包含丰富的边缘能量信息, 能够捕捉水墨画局部细节运笔信息和典型笔触的纹理特征, 本文利用 Daubechies 第 1 层小波系数矩阵作为纹理输入. 对于构图, 通过计算水墨画的显著图来分析画面中各个元素的空间分布结构, 采用 SLIC 超像素分割方法^[29]将原图像分割成若干均匀区域, 如图 6a 所示. 之后采用 Cheng 等^[30]提出的基于直方图对比的方法来计算灰度显著图, 如图 6b 所示, 其中较亮区域对应显著度值越高, 其审美视觉上越重要.

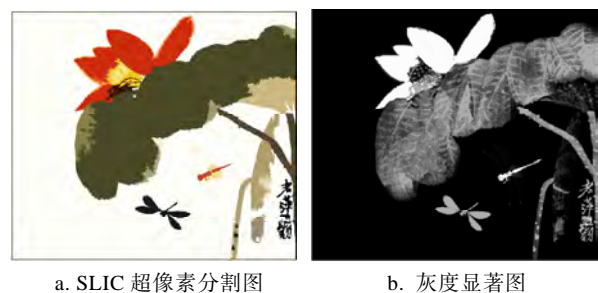


图 6 水墨画显著图提取

对于局部视角输入, 将水墨画表示为裁剪图像块的集合. 与随机裁剪方法^[31]不同, 本文采用 Ma 等^[19]提出的自适应策略来提取水墨画中审美感知信息最多的图像块, 其准则有如下几点: (1) 显著检测能够突出画面中的视觉重要区域, 因此使用显著图去选择图像中最有辨识性和信息量的区

域; (2) 水墨画中重要的审美细节都要通过不同主体与背景的关系来表现, 所以图像块内要保证模式多样性; (3) 约束不同图像块之间的空间距离要进行约束, 从而以保证图像块之间的重叠率尽可能小. 因此, 将水墨画图像块选择作为一个优化问题, 定义目标函数^[19]

$$\{c^*\} = \arg \max F(S, D_p, D_s) \quad (2)$$

$$F(\cdot) = \sum_{i=1}^M S_i + \sum_{i \neq j}^M D_p(\tilde{N}_i, \tilde{N}_j) + \sum_{i \neq j}^M D_s(c_i, c_j) \quad (3)$$

其中, $\{c_m^*\}_{m \in [1, M]}$ 表示最优化图像块集合每个中心坐标; M 表示图像块数量; S_i 表示每个图像块 p_i 的归一化显著度值; $D_p(\cdot)$ 和 $D_s(\cdot)$ 分别表示模式距离函数以及欧几里得空间距离函数. 本文采用多元高斯对每个图像块 p_m 模式建模^[19], 即

$$\tilde{N}_m = \{\{N_e(\mu_e, \Sigma_e)\}_m, \{N_c(\mu_c, \Sigma_c)\}_m\}_{m \in [1, M]} \quad (4)$$

其中, $\{N_e(\mu_e, \Sigma_e)\}_m$ 和 $\{N_c(\mu_c, \Sigma_c)\}_m$ 分别表示该图像块的笔触边缘分布和浓淡色度分布, 其中的模式距离用来衡量 2 个图像块的差异性. 图 7 展示了若干采用自适应选择策略提取的图像块结果, 其大小均为 $224 \times 224 \times 3$. 该方法不仅可以有效地选出最显著区域(如花朵、枇杷和鸟), 而且能捕捉不同水墨画主体与留白背景之间的模式多样性(如树干与山水、枝叶与草虫).



图 7 自适应图像块选择策略提取的图像块示例

4.2.2 统计聚合结构

本文采用 Lu 等^[31]提出的统计聚合结构来组合各路经过 VGG16 网络输出的特征. 以局部图像块为例, 用 $D = \{d_m\}_{m \in [1, M]}$ 表示由 VGG16 网络组输出的图像块特征集合, 这里 d_m 是 K 维向量, $T_m = \{d_{mk}\}_{m \in [1, M]}$ 用来表示所有 $d_m \in D$ 第 k 分量集合, \oplus 用来表示向量级联操作. 统计层由统计函数集合 $F_{\text{Agg}} = \{F_{\text{Agg}}^u\}_{u \in [1, U]}$ 构成, $U = \{F_{\text{min}}, F_{\text{max}}\}$. 该集合的输出级联后由全连接层聚合产生一个 K_{stat} 维特征向量. 整个结构函数^[31]表示为

$$g(D) = W \times (\oplus_{u=1}^U \oplus_{k=1}^K F_{\text{Agg}}^u(T_k)) \quad (5)$$

其中, $W \in \mathbb{R}^{K_{\text{stat}} \times UK}$ 为全连接层参数, 图 8 所示为 $M = 4, K = 3$ 的统计聚合结构图, 在实验中设置 $M = 5, K = 256$.

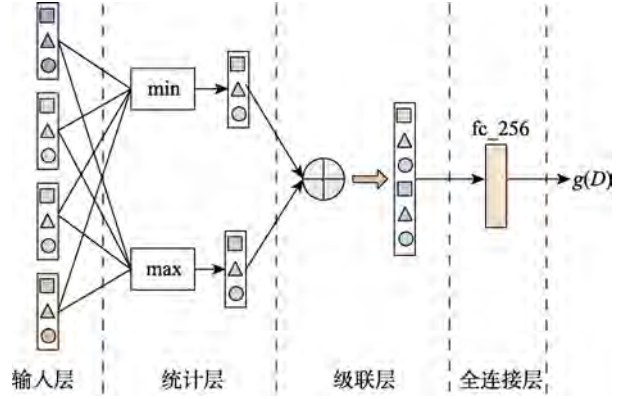


图 8 统计聚合结构图^[31]

4.2.3 训练细节

首先对各路基准网络进行单独微调, 以此作为多视角并行网络的初始化权重; 然后分别对全局视角和局部视角 2 个子聚合网络进行训练, 其中全局输入为 6 个通道, 局部输入为 5 个通道; 最后将 2 个子网络输出的 256 维向量进行级联, 并对回归层权重进行训练, 采用最后输出的 512 维向量作为水墨画的深度审美特征表示.

5 自适应水墨画深度审美评估模型

不同题材类型的水墨画作品对应着不同的审美评估标准, 比如山水画侧重于笔墨技法的皴染点擦, 而花鸟画侧重于色彩点染. 因此, 本节利用第 4 节提取的深度审美特征, 采用 Tian 等^[16]提出的照片查询机制, 提出一种基于题材查询的自适应水墨画审美评估模型, 即根据测试图像在训练集中找到视觉内容或语义相似的图像集合, 进而训练相应的模型来进行审美预测.

具体地, 每给定一幅测试图像 I_q , 首先提取它的视觉表示特征, 本文使用 AlexNet 在 ImageNet 上的预训练模型, 并使用水墨画训练数据微调, 去除原网络最后一个包含 1000 个神经元的全连接层, 添加上包含 6 个神经元的全连接层作为分类结果, 用于代表输出水墨画 6 类题材, 以此适应水墨画题材分类任务. 然后提取微调后模型的第 7 层全连接特征 $x_I = [x_1, x_2, \dots, x_{4096}]^T$ 来量化表示水墨画的题材内容. 这种情况下, 测试图片 I_q 与训练集

中图片 I_d 间的相似度可计算公式为

$$\text{Sim}_{\text{visual-CNN}}(I_q, I_d) = 1 - \frac{1}{2} \left\| \frac{x_q}{\|x_q\|_1} - \frac{x_d}{\|x_d\|_1} \right\|_1 \quad (6)$$

之后利用该相似度函数在整个训练集中查找它的相似图片. 利用这种基于题材检索的返回结果, 可以构建相应的自适应子训练集

$$L^q = \{I_d | I_d \in L \cap I_d \in N(I_d)\} \quad (7)$$



图 9 使用 QDep_IR_{CNN} 机制构造的水墨画自适应训练集

6 实验结果与性能分析

本文所有的网络训练和测试采用 Tensorflow 深度学习框架^[32]完成, 初始学习率为 0.001, 每 10 轮缩小 10 倍, 权重衰减率为 10^{-5} , 动量为 0.9. 本文将 1200 幅作品进行水平映射变化扩大为 2400 幅, 之后将所有作品随机分为 5 组, 每组 480 幅素描作品; 采用 5 折交叉验证法, 其中 4 组为训练样本, 1 组为测试样本, 使用随机梯度下降法训练, 5 次实验后的均值为模型最终结果. 在模型性能评估上, 选取测试集并计算审美模型预测分数与人工评估分数的平均皮尔森相关系数 (R_p / Sig.) 与均方误差 (mean squared error, MSE); 其中, Sig. 即 p 值, 代表假设检验中的显著性, R_p 越高且 Sig. < 0.05 时, 表示该模型在统计意义上具有显著性相关; 并且 MSE 越小, 表示预测误差越小, 其模型性能越好. 本节将比较不同基础架构和多视角输入的并行深度 CNN 性能, 并分析基于题材查询的自适应审美模型的预测性能.

6.1 不同架构和视角输入下模型性能比较

首先以水墨画原图像作为单视角输入, 比较

其中, L 为训练集; $N(I_d)$ 为根据视觉表示特征计算的与测试图像最相似的图片集合, 检索过程表示为 QDep_IR_{CNN}. 若采用 QDep_IR_{CNN} 机制构造的水墨画自适应子训练集如图 9 所示. 接着在 L^q 上利用前面多视角并行网络输出的深度审美特征, 使用支持向量回归 (support vector regression, SVR) 训练得到自适应审美评估模型, 并利用该模型预测得到该图像的审美感知分数.

不同基础架构下单路深度 CNN 模型 (single-column deep convolutional neural networks, SCNN) 对水墨画的审美评估性能, 如表 2 所示. 其中, Arc1 表示原始 VGG16 网络结构, Arc2 表示仅减少了 2 个全连接层神经元数量而无题材层的 VGG16 网络结构, Arc3 表示将第 13 个卷积层修改为 6 个并行题材卷积组后的 VGG16 网络结构 (本文模型采用的基础架构). 可以看出, Arc3 架构下的模型相较于其他架构有较高的平均 R_p (Sig. < 0.05) 和较低的平均 MSE, 因而对水墨画有较高的审美评估性能. 这说明在较少网络参数的前提下, 本文网络架构和迁移学习策略既可以充分地利用预训练模型浅层的边缘颜色等相关通用审美特征, 又能够结合自适应的题材卷积组和回归损失层来有效地捕捉不同水墨画题材的专业审美要素.

表 2 不同基础架构下 SCNN 模型性能比较

架构	R_p /Sig.	平均 MSE
Arc1 conv13(512)_fc4096_fc4096	0.674/0.000	0.336
Arc2 conv13(512)_fc512_fc256	0.661/0.000	0.348
Arc3 conv13(64)_fc512_fc256	0.712/0.000	0.285

注: 粗体表示本文模型的基础架构及模型性能.

在此基础上, 本节比较不同视角输入下的并行深度 CNN(multi-view deep convolutional neural networks, MVCNN)模型对水墨画审美感知的预测性能, 并与之前传统方法进行比较, 如表 3 所示. 为了验证深度学习方法的有效性, 将 MVCNN 与一些使用人工设计特征的模型, 如 AVA^[4]和张佳婧等^[13]提出的线性回归模型作为基准进行性能比较. 可以看出, 所有深度学习特征的审美评估性能要显著优于手工设计特征, 从而验证了深度学习对水墨画审美量化评估的有效性. 这种优越的表现源自深度神经网络能基于人工评分训练数据直接从原始水墨画中提取高层审美语义特征.

表 3 不同视角输入下 MVCNN 与传统方法性能比较

方法	R_p /Sig.	平均 MSE
AVA ^[7] (2012)	0.462/0.000	0.532
张佳婧等 ^[13] (2017)	0.556/0.000	0.446
RAPID ^[15] (2014)	0.645/0.000	0.371
DMA-Net ^[31] (2015)	0.654/0.000	0.352
MSDLM ^[17] (2016)	0.669/0.000	0.344
SCNN-Arc3	0.712/0.000	0.285
MVCNN-global	0.731/0.000	0.267
MVCNN-local	0.782/0.000	0.212
MVCNN-hybrid	0.791/0.000	0.193

注: 粗体表示本文融合全局和局部视角的并行 CNN 模型以及预测性能.

此外, 本节还将 MVCNN 模型与一些已有的照片审美评估模型例如 RAPID^[15], DMA-Net^[31]和 MSDLM^[17]进行比较, 结果如表 3 所示. 可以看出, MVCNN 模型在水墨画审美预测方面显著优于上述方法. 该结果进一步验证了 MVCNN 模型的有效性, 特别是自适应选择策略能够有效地提取水墨画中审美感知信息最多的显著区域, 而且能捕捉不同水墨画主体与留白背景之间的模式多样性, 例如, 笔触与留白形成的虚实和疏密对比、色墨对比、动静对比等.

最后, 本节比较在 Arc3 架构下不同视角输入下的模型预测性能. 其中, SCNN-Arc3 表示将原图像缩放到 $224 \times 224 \times 3$ 固定大小作为单路输入的 VGG16 模型, MVCNN-global 和 MVCNN-local 分别表示以全局图像和自适应局部图像块作为输入的并行网络模型, 而 MVCNN-hybrid 表示融合全局与局部视角输入的并行网络模型. 实验结果表明, 3 种多路并行网络模型 MVCNN-global, MVCNN-local 和 MVCNN-hybrid 的预测性能要优于单路输入的 SCNN-Arc3 模型, 而 MVCNN-hybrid 模型相较于其

他单视角模型有较高的平均 R_p (Sig.<0.05) 和较低的平均 MSE, 因而在水墨画审美评估方面有较高性能. 这表明模型能够从整体布局信息和局部精细细节 2 个视角综合提取水墨画审美特征.

6.2 自适应深度审美模型性能分析

为了进一步验证自适应模型(第 6 节)预测性能, 将检索过程 QDep_IR_{CNN} 中的自适应子训练集大小设为 50, 并使用 SVR 将第 5 节 MVCNN-hybrid 模型提取的深度审美特征与人工评估分数进行训练, 其得到的审美评估模型表示 IR_{CNN}+SVR_{DeepAesth}. 该模型与 MVCNN-hybrid 模型性能进行比较, 如表 4 所示. 可以看出 IR_{CNN}+SVR_{DeepAesth} 比 MVCNN-hybrid 有较高的 R_p 和较低的平均 Sig.<0.05, 其模型评估结果与人工审美评分之间达到 0.823 的皮尔森高度显著相关. 这表明自适应模型能够有效地捕捉水墨画不同题材内容对审美标准的影响, 而模型中的深度审美特征在水墨画计算美学评估方面具有较高的预测性能.

表 4 自适应审美模型与 MVCNN-hybrid 性能比较

方法	R_p /Sig.	平均 MSE
MVCNN-hybrid	0.791/0.000	0.193
IR_{CNN}+SVR_{DeepAesth}	0.823/0.000	0.161

注: 粗体表示本文自适应深度审美模型以及预测性能.

一些在审美感知评估分数上呈现对立方向的水墨画图像比较如图 10 所示. 其中, 顶行作品美感分数较高, 而底行作品美感分数较低. 图中的数值表明, 自适应深度审美模型可以很好地描述和预测水墨画中的美感属性, 并且与人类审美感知相吻合. 此外, 顶行的水墨画图像与底行相比墨色浓淡层次变化更加自然, 笔触更加顺畅有序, 干湿对比更加和谐, 并且在画面整体布局上空间感更好. 由于不同艺术家以不同表现形式灵活地应用审美准则, 从而形成风格迥异的水墨画艺术. 这些作品与数据集中大多数相应题材的一般审美准则相比有较大偏差, 容易被模型评价为美感分数较低, 但这些著名国画大师的作品因其极高的艺术价值被人工评价为美感分数较高. 风格抽象的水墨画表现形式是导致水墨画作品美学评价有较大误判的主要因素, 图 11 展示了若干评价分数偏差较大的样本.

此外, 本节进一步比较了基于人工设计特征的线性回归模型等^[13]、多视角并行深度学习模型 MVCNN-hybrid 和基于题材查询的自适应深度审



图 10 人工(蓝色)和模型(红色)评估分数比较



图 11 较大误判的人工(蓝色)和模型(红色)评估分数比较

美模型 $IR_{CNN}+SVR_{DeepAesth}$ 在 6 类水墨画题材的审美预测性能, 如图 12 所示. 图中显示 $IR_{CNN}+SVR_{DeepAesth}$ 在 6 个类别中显著且持续优于其他 2 种模型, 从而进一步证明通过多视角并行 CNN 学习的审美特征的有效性.

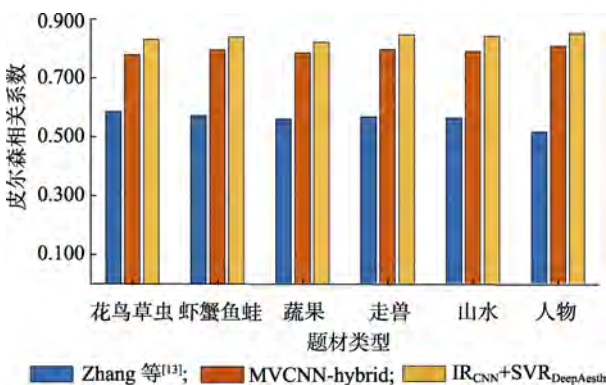


图 12 对不同水墨画题材类别的审美预测性能比较

6.3 深度审美模型敏感度分析

为了进一步揭示深度审美模型的可解释性, 这里对构图、墨色、纹理 3 个重要的国画绘画要素进行实验干扰, 并测试深度 CNN 响应对这些因素

变化的敏感度.

首先选取 100 个大小从 10×10 到 50×50 像素的小方块, 并在水墨画图像上任意位置进行随机遮挡, 从而对整体布局进行干扰, 如图 13 所示. 对应的性能结果如图 14 中的蓝色线条所示. 可以看出其模型预测结果与人工评分之间的相关系数随着布局干扰的增大而快速衰减, 这表明深度审美模型对空间布局较为敏感.

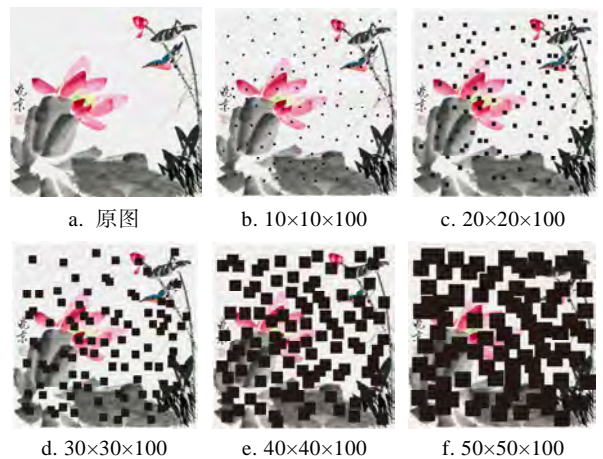


图 13 对不同大小方块随机遮挡下的水墨画图像

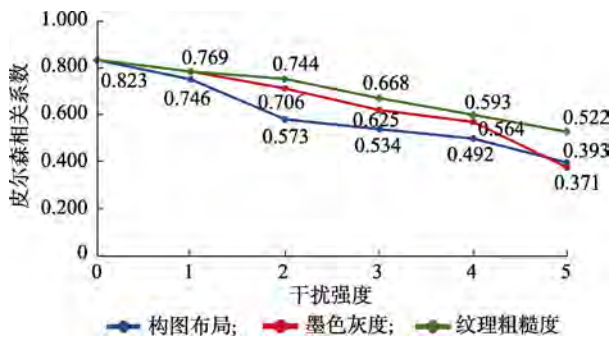


图 14 不同干扰强度下皮尔森相关系数变化

其次, 设置不同的灰度系数 g 来对水墨画图像整体色阶进行干扰, 其范围为 $0\sim 1$, 越接近 0 , 图像越暗, 如图 15 所示; 对应的性能结果如图 14 中的红色线条所示. 可以看出, 随着灰度系数的下降, 其相关系数随着色阶干扰的增大而逐渐下降, 表明模型对墨色灰度的层次变化较为敏感.

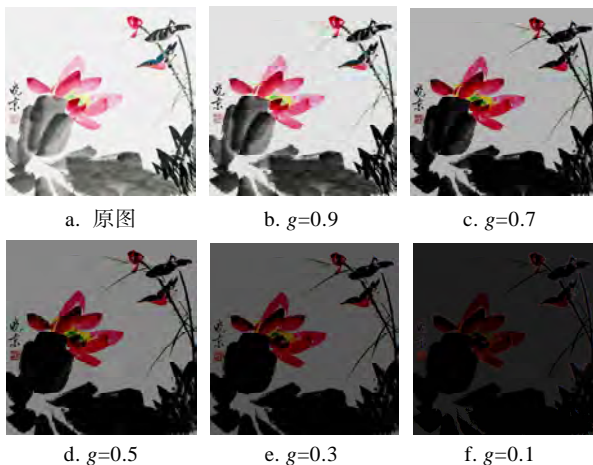


图 15 不同灰度系数对色阶干扰下的水墨画图像

最后, 对水墨画图像添加不同程度的高斯噪声来对纹理进行干扰, 其方差参数 s 越大, 图像越粗糙, 如图 16 所示; 对应的性能结果如图 14 的绿

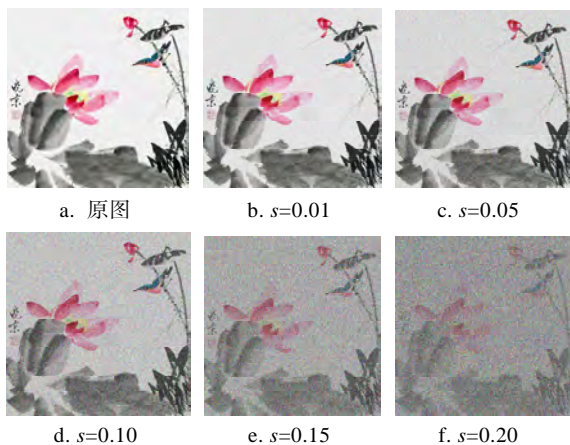


图 16 不同方差噪声对纹理干扰下的水墨画图像

色线条所示. 可以看出, 随着方差的增大, 其相关系数随着噪声增大而逐渐下降, 表明深度审美模型对笔触纹理较为敏感.

7 结论和未来工作

本文以水墨画作为对象, 提出一种基于深度学习的自适应计算美学评估框架. 实验结果表明, 自适应深度审美模型评估结果与人工审美评估结果之间达到 0.823 的皮尔森高度显著相关, 其性能明显优于传统的基于手工设计特征模型, 证明了模型的有效性, 并且通过多视角并行 CNN 学习到的深度审美特征能够准确和敏锐捕捉构图、墨色、纹理 3 种绘画要素的变化. 该工作为水墨画作品定量审美评估提供一个基于深度学习技术的参考框架, 它在一个新的维度上加深了对国画作品审美感知的理解, 同时为其他绘画艺术形式的计算美学问题提供新的思路与解决方案.

在未来工作中, 将继续探索将手工设计特征与深度学习特征结合来提高模型对水墨画的审美评估性能. 此外, 将深度 CNN 扩展到水墨画的多任务深度学习上, 研究通过水墨画其他高层语义例如情感意境的预测来更好地辅助水墨画审美评估任务.

参考文献(References):

- [1] Hoenig F. Defining computational aesthetics[C] //Proceedings of the 1st Eurographics Conference on Computational Aesthetics in Graphics, Visualization and Imaging. Aire-la-Ville: Eurographics Association Press, 2005: 13-18
- [2] Fishwick P A. Aesthetic computing[M]. Cambridge: MIT Press, 2006
- [3] Luo W, Wang X G, Tang X O. Content-based photo quality assessment[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2011: 2206-2213
- [4] Murray N, Marchesotti L, Perronnin F. AVA: a large-scale database for aesthetic visual analysis[C] //Proceedings of the 25th IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2408-2415
- [5] Sartori A, Yanulevskaya V, Salah A A, et al. Affective analysis of professional and amateur abstract paintings using statistical analysis and art theory[J]. ACM Transactions on Interactive Intelligent Systems, 2015, 5(2): Article No.8
- [6] Amirshahi S A, Hayn-Leichsenring G U, Denzler J, et al. JenAesthetics subjective dataset: analyzing paintings by subjective scores[C] //Proceedings of the European Conference on

- Computer Vision. Heidelberg: Springer, 2015: 3-19
- [7] Ke Y, Tang X O, Jing F. The design of high-level features for photo quality assessment[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2006: 419-426
- [8] Wang Chaohui, Pu Yuanyuan, Xu Dan, *et al.* Evaluating aesthetics quality in portrait photos[J]. *Journal of Software*, 2015, 26(S2): 20-28(in Chinese)
(王朝晖, 普园媛, 徐丹, 等. 人像照片的美感质量评价[J]. *软件学报*, 2015, 26(S2): 20-28)
- [9] Dhar S, Ordonez V, Berg T L. High level describable attributes for predicting aesthetics and interestingness[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2011: 1657-1664
- [10] Obrador P, Saad M A, Suryanarayan P, *et al.* Towards category-based aesthetic models of photographs[C] //Proceedings of the 18th International Conference on Multimedia Modeling. Heidelberg: Springer, 2012: 63-76
- [11] Su H H, Chen T W, Kao C C, *et al.* Preference-aware view recommendation system for scenic photos based on bag-of-aesthetics-preserving features[J]. *IEEE Transactions on Multimedia*, 2012, 14(3): 833-843
- [12] Li C C, Chen T. Aesthetic visual quality assessment of paintings [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2009, 3(2): 236-252
- [13] Zhang Jiajing, Peng Ren, Wang Jian, *et al.* Computational aesthetic evaluation of Chinese wash paintings[J]. *Journal of Software*, 2016, 27(S2): 220-233(in Chinese)
(张佳婧, 彭韧, 王健, 等. 水墨画计算审美评估[J]. *软件学报*, 2016, 27(S2): 220-233)
- [14] Donahue J, Jia Y G, Vinyals O, *et al.* DeCAF: a deep convolutional activation feature for generic visual recognition[C] //Proceedings of the 31st International Conference on International Conference on Machine Learning. New York: ACM Press, 2014, 32: I-647-I-655
- [15] Lu X, Lin Z, Jin H, *et al.* RAPID: rating pictorial aesthetics using deep learning[C] //Proceedings of the 22nd ACM International Conference on Multimedia. New York: ACM Press, 2014, 457-466
- [16] Tian X M, Dong Z, Yang K Y, *et al.* Query-dependent aesthetic model with deep learning for photo quality assessment[J]. *IEEE Transactions on Multimedia*, 2015, 17(11): 2035-2048
- [17] Wang W N, Zhao M Q, Wang L, *et al.* A multi-scene deep learning model for image aesthetic evaluation[J]. *Signal Processing: Image Communication*, 2016, 47: 511-518
- [18] Kao Y Y, He R, Huang K Q. Deep aesthetic quality assessment with semantic information[J]. *IEEE Transactions on Image Processing*, 2017, 26(3): 1482-1495
- [19] Ma S, Liu J, Chen C W. A-Lamp: adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 722-731
- [20] Talebi H, Milanfar P. NIMA: neural image assessment[J]. *IEEE Transactions on Image Processing*, 2018, 27(8): 3998-4011
- [21] Michal K, Alexander C L, David W M. Leveraging expert feature knowledge for predicting image aesthetics[J]. *IEEE Transactions on Image Processing*, 2018: 27(10): 5100-5112
- [22] Li Chao, Sun Shouqian, Min Xin, *et al.* Application of deep convolutional features in sketch works classification and evaluation[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2017, 27(10): 1898-1904(in Chinese)
(李超, 孙守迁, 闵歆, 等. 深度卷积特征在素描作品分类与评价中的应用[J]. *计算机辅助设计与图形学学报*, 2017, 27(10): 1898-1904)
- [23] Fan Z B, Zheng X S, Zhang K. Computational analysis and eye movement experiments of white space in Chinese paintings[C] //Proceedings of the IEEE International Conference on Progress in Informatics and Computing. Los Alamitos: IEEE Computer Society Press, 2015: 301-306
- [24] Yokochi S, Okada T. Creative cognitive process of art making: a field study of a traditional Chinese ink painter[J]. *Creativity Research Journal*, 2005, 17(2/3): 241-255
- [25] Rawson J. *The British museum book of Chinese art*[M]. London: The British Museum Press, 2007
- [26] Cui Nianhan. Exquisite works of vivid charm-interpretation of the "First of the Six Laws" in Xiehe's collection of ancient paintings[J]. *Beauty and Age*, 2016, 649(4): 38-39(in Chinese)
(崔念哈. 风神逸品, 气韵生动——解读谢赫《古画品录》中的“六法之首” [J]. *美与时代*, 2016, 649(4): 38-39)
- [27] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C] //Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: ACM Press, 2012: 1097-1105
- [28] Yosinski J, Clune J, Bengio Y, *et al.* How transferable are features in deep neural networks[J]. *Advances in Neural Information Processing Systems*, 2014, 27: 3320-3328
- [29] Perazzi F, Krähenbühl P, Pritch Y, *et al.* Saliency filters: contrast based filtering for salient region detection[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 733-740
- [30] Cheng M M, Mitra N J, Huang X, *et al.* Global contrast based salient region detection[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(3): 569-582
- [31] Lu X, Lin Z, Shen X, *et al.* Deep multi-patch aggregation network for image style, aesthetics, and quality estimation[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 990-998
- [32] Abadi M, Barham P, Chen J, *et al.* TensorFlow: a system for large-scale machine learning[C] //Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. New York: ACM Press, 2016: 265-283