

基于并行深度卷积神经网络的图像美感分类

王伟凝¹ 王励¹ 赵明权¹ 蔡成加¹ 师婷婷¹ 徐向民¹

摘 要 随着计算机和社交网络的飞速发展, 图像美感的自动评价产生了越来越大的需求并受到了广泛关注. 由于图像美感评价的主观性和复杂性, 传统的手工特征和局部特征方法难以全面表征图像的美感特点, 并准确量化或建模. 本文提出一种并行深度卷积神经网络的图像美感分类方法, 从同一图像的不同角度出发, 利用深度学习网络自动完成特征学习, 得到更为全面的图像美感特征描述; 然后利用支持向量机训练特征并建立分类器, 实现图像美感分类. 通过在两个主流的图像美感数据库上的实验显示, 本文方法与目前已有的其他算法对比, 获得了更好的分类准确率.

关键词 图像美感评估, 深度卷积神经网络, 并行卷积神经网络, 特征提取

引用格式 王伟凝, 王励, 赵明权, 蔡成加, 师婷婷, 徐向民. 基于并行深度卷积神经网络的图像美感分类. 自动化学报, 2016, 42(6): 904–914

DOI 10.16383/j.aas.2016.c150718

Image Aesthetic Classification Using Parallel Deep Convolutional Neural Networks

WANG Wei-Ning¹ WANG Li¹ ZHAO Ming-Quan¹ CAI Cheng-Jia¹ SHI Ting-Ting¹ XU Xiang-Min¹

Abstract With the rapid development of computers and social networks, automatic image aesthetic evaluation is in demand and has attracted more and more attention recently. Since the complexity and subjectivity of image aesthetic evaluation task, the traditional handcrafted features and generic image descriptors are hard to represent the overall aesthetic character of images. It is difficult for them to quantify and model the image aesthetics exactly. In this paper, a new method of image classification based on parallel deep convolutional neural networks is proposed. We use parallel deep learning networks to automatically complete feature extraction and acquire more comprehensive description of image aesthetics from different views. Then a support vector machine (SVM) classifier is built with the aesthetic features to accomplish image aesthetic classification. Experiments on two most frequently used databases of image aesthetics demonstrate that our proposed method achieves better results than other existing methods.

Key words Image aesthetic assessment, deep convolutional neural networks, parallel convolutional neural networks, features extraction

Citation Wang Wei-Ning, Wang Li, Zhao Ming-Quan, Cai Cheng-Jia, Shi Ting-Ting, Xu Xiang-Min. Image aesthetic classification using parallel deep convolutional neural networks. *Acta Automatica Sinica*, 2016, 42(6): 904–914

近年来, 随着社交网络的普及和发展, 数字图像数量激增. 每天被上传和分享的图像都不计其数, 大规模图像的自动管理需求越来越大. 人们往往倾向于将高质量的照片留存下来. 近些年, 从美感的角度来评价图像质量受到了越来越广泛的关注^[1]. 目前,

图像的自动美感评估已有许多应用. 例如, 在图像检索方面, 美感质量成为排序算法的重要参考之一; 在图像美学优化方面, 美感评价具有重要的指导意义^[2]; 在人机交互方面, 图像美感有利于建立和谐的人机交互系统^[3–4].

美学是以艺术为主要对象, 研究美、丑等审美范畴和人的审美意识、美感经验、美的创造、发展及其规律的科学^[1]. 图像美感可以定义为人们在观察图像时的美学感受^[3]. 美感评估是希望通过计算机, 模拟人类的感知来判断图像的美感, 实现对图像进行美感高低分类, 或者对图像的美感程度给出评分. 在计算机视觉领域, 由于人类美感活动的复杂性和主观性, 美感质量评估研究具有相当的挑战性. 美感知受到文化、时代、地域、个人主观喜好等的影响. 但总的说来, 人类的美感感受和判断还是具有很大的共性. 无数绘画、摄影和艺术作品作为人类共同的

收稿日期 2015-10-31 录用日期 2016-02-27
Manuscript received October 31, 2015; accepted February 27, 2016

国家自然科学基金 (61171142, 61401161), 广东省自然科学基金 (2015A030313212), 广东省前沿与关键技术创新专项资金 (重大科技专项) (2014B010111003, 2014B010111006) 资助

Supported by National Natural Science Foundation of China (61171142, 61401161), Natural Science Foundation of Guangdong Province (2015A030313212), and Guangdong Frontier and Key Technological Innovation Special Funds (Grant Scientific and Technological Project) (2014B010111003, 2014B010111006)

本文责任编辑 柯登峰

Recommended by Associate Editor KE Deng-Feng

1. 华南理工大学电子与信息学院 广州 510641

1. School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641

审美财富受到人们的广泛欣赏和喜爱。

在图像美感评价研究中, 早期研究者们通过提取图像的相关美学特征, 再利用机器学习方法来训练分类器, 对图像进行高低美感分类。美学特征的有效提取对分类结果至关重要。研究者们对美学特征做了大量的研究, 从人类的审美习惯和经验出发, 借鉴摄影、艺术、绘画等领域的美感规则, 考虑人类的视觉注意机制等, 从图像中提取了各类美学特征^[4-9]。Tong 等^[5] 率先提取了许多低层的基本图像特征, 如颜色直方图、图像能量等。然而这些特征缺乏美感相关性。Datta 等^[6] 则提取了 56 维包括与图像美感较为相关的低层特征和基于美感规则的高层特征 (如景深、三分法则等)。Ke 等^[8] 也设计了高层美学特征, 用更少的特征数取得了更好的效果。通过对主体区域的研究, Tang 等^[9] 利用区域特征和全局特征来进行图像美感评价, 取得了很好的效果。这些手工特征在图像美感评价中都取得了一定的效果, 促进了图像美感评价的发展进步。然而, 图像的构图规则和美学规律千变万化, 在工程上难以被准确地建模和量化; 人工提取的美学特征往往是从图像的一个小角度出发, 难以全面地表达图像美感; 且一些高层美学特征往往只是对美学规则的近似^[10]。这些都已成为图像美学特征提取中的瓶颈。

为了克服传统手工特征 (Handcrafted features) 设计存在的问题, 一些研究者尝试用局部特征 (Generic features)^[11-12] 来评估图像美感质量。这些特征能够详细描述图像的局部信息和细节, 例如 Marchesotti 等^[11] 利用 Bag-of-Visual-Words 局部特征; Guo 等^[12] 对图像的 SIFT 特征进行 LLC (Locality-constrained linear coding) 编码, 并结合手工特征, 都取得了一定的效果。然而, 局部特征应用于图像美感分析时, 存在较大的局限性。例如, 它忽略了图像的色彩信息, 而图像色彩对其美感非常重要; 同时也忽略了全局信息, 因此难以从整体的构图或美感规律上来表达图像。另外, 各类局部特征大都是提取 SIFT 特征后通过不同的编码得到, 特征设计的思路没有本质性的突破, 难以较大地提升分类准确率。

近几年, 深度学习方法的出现, 在机器学习和特征学习方面引发了一场革命。与传统的机器学习方法不同, 深度学习无需人工提取特征, 其网络可以根据输入自动学习特征, 在解决许多传统人工智能, 包括计算机视觉问题上有突破性进展, 如图像分类、物体识别、人脸识别等^[13-15]。其在物体分类任务中带来的准确率的显著提升是深度卷积神经网络 (Convolutional neural networks, CNN) 获得人们的广泛关注的里程碑。深度学习网络利用大量的原始样本数据, 充分利用隐藏层逐层深入地学习图像

的抽象信息, 更全面直接地获取图像特征。通过对特征的分析 and 从其卓越的效果可以看出, 深度网络中学习的特征能更好地表达图像的不同类别信息。数字图像是以矩阵来描述的, 卷积神经网络能更好地从局部信息块出发, 进而描述图像的整体结构, 故在计算机视觉领域, 深度学习方法中大多采用卷积神经网络来解决问题。

2014 年以来, 学者们开始尝试将深度学习用于图像美感分类问题。Lu 等^[10] 同时考虑了图像的局部视角和全局视角来设计卷积神经网络, 并将图像的风格和语义标签用于美感质量评价。Dong 等^[16] 利用大型图像物体分类与检测数据库训练得到的深度学习网络模型来提取图像特征, 基于这些特征训练一个两类支持向量机 (Support vector machine, SVM) 分类器, 对图像进行高低美感分类。另一方面, Dong 等^[17] 还进一步将上面的深度学习网络特征和手工特征进行了特征融合, 用于 SVM 分类器, 取得了较好的效果。总的来说, 利用深度学习网络解决图像美感评估问题还只是刚刚起步, 还有很多内容值得研究探索和发展, 分类效果也有待进一步提高。

本文尝试使用深度学习方法进行图像美感评估。我们认为在组建图像美感分类问题的深度学习网络时, 有以下三个问题值得考虑:

1) 深度学习网络能否学习出图像美学特征?

现有传统方法提取的美学特征在不同程度上存在有效性低、冗余性大、鲁棒性低的问题, 图像美感分类的效果受到限制^[17]。深度学习网络是模拟生物视觉系统进行设计的, 具有较强的图像理解能力。

深度学习网络在图像物体分类问题上取得了很好的效果。其可以从原始图像中自动学习出包含丰富语义信息的特征矩阵, 这些信息很好地表达了图像中的物体类别信息。从已有图像物体分类的研究成果来看^[13], 深度学习网络自动学习到的特征, 比手工特征或局部特征的分类准确率更高。

图像美感分类和图像物体分类都是对图像矩阵进行分析, 从中获取图像类别间共性和个性的特征, 以求最大化地区分不同类别。我们认为, 深度网络提取的图像信息中, 必定也包含了图像美感质量的内容, 利用深度学习网络学习美学特征, 是图像美感评估的一个发展方向。

2) 如何解决深度学习网络的学习能力与图像美感数据库数据量偏小的矛盾?

已有的深度学习应用研究指出^[14], 深度学习网络适用于解决数据量大的复杂问题, 例如 ILSVRC 竞赛中的图像分类问题, 其是一个包含 120 万幅图像 1000 个物体类别的分类问题, 图像数量大, 类别多, 采用卷积神经网络能比传统方法取得更好的效

向量, 即我们要提取的特征向量. 网络具体结构设计思路见第 2.1 节.

并行网络的输入分别为: 1) 原图像缩放到宽和高都为 256 像素的图像矩阵; 2) 图像的 HSV 变换的 V 通道矩阵; 具体介绍见第 2.2 节.

将训练集中的图像分别转换为以上两种输入形式, 输入到单路卷积神经网络中训练得到网络模型. 然后对每幅图像分别利用每一路的网络模型提取第 5 个全连接层的特征, 并将特征结合起来, 则每幅图像得到一个 512 维的特征. 利用支持向量机 (SVM) 对这些特征进行训练, 得到最终的图像美感分类器.

在训练完成后, 系统对于一幅新图像的输入, 先按上述两种输入形式做变换, 得到网络的输入矩阵. 然后分别利用每一路卷积神经网络的模型提取图像特征, 将特征合并得到 512 维特征, 再经过图像美感分类器进行分类, 自动得到高或低美感评价结果.

2 并行深度卷积神经网络实现

2.1 单路卷积神经网络设计

针对图像美感数据库图像数据少和复杂性低的问题, 我们将大型图像数据库训练出来的网络模型的一部分, 迁移到我们的美感评价网络中, 以其训练所得的参数作为我们网络的参数初始值, 以使网络能在已包含图像基本信息的基础上学习美学质量信息.

深度卷积神经网络方法由 Krizhevsky 等^[13] 于 2012 年首次提出, 其在一年一度著名的 ImageNet 大规模计算机视觉识别挑战赛 (ImageNet Large Scale Visual Recognition Challenge, ILSVRC) 的大型图像分类数据库上的分类效果上取得了非常大的进步, 具有里程碑式的意义. 这个图像分类数据库中的图像来自于目前世界上图像识别最大的数据库 ImageNet, 根据图像中物体种类的不同分为 1000

个类, 包含约 120 万幅图像. 这样的复杂任务使得训练出来的深度学习网络可以学习到足够多的图像信息来更好地解析表达图像.

美感分类与图像分类问题具有较好的相关性, 已有的图像美感评估研究显示^[19], 在基于内容的图像检索和分类中有效的一些低层特征或者语义特征, 对于图像美感分类都发挥了较好的作用. 因此我们可以合理地假设已完成 ImageNet 数据库训练的图像物体分类网络模型^[13], 具有提取复杂图像特征的能力, 可以作为图像美学特征提取的基础. 并且, 文献 [13] 的网络是目前已有图像分类任务模型中较为简洁的一个网络结构, 为我们组建并拓展美感深度学习网络提供了便利.

本文的卷积神经网络结构的设计是在文献 [13] 网络结构的基础上, 进行了适应性的扩展. 我们采用文献 [13] 网络结构的前 6 层, 并就本文的算法思路添加了 5 层全连接层. 具体如下:

本文网络的前 6 层具体为: 第 1 层为卷积层, 有 96 个卷积核, 大小为 $11 \times 11 \times 3$; 第 2 层为卷积层, 有 256 个卷积核, 大小为 $5 \times 5 \times 48$; 第 3 层为卷积层, 有 384 个核, 大小为 $3 \times 3 \times 256$; 第 4 层为卷积层, 有 384 个核, 大小为 $3 \times 3 \times 192$; 第 5 层为卷积层, 有 256 个核, 大小为 $3 \times 3 \times 192$; 第 6 层为全连接层, 有 4096 个输出. 这些卷积层间的池化层和归一化层的设置, 以及层间的连接与文献 [13] 的结构相同.

为了适应本文的算法思路, 我们对单路卷积神经网络接下来的全连接层进行了探索和设计. 由于后面要对特征进行联合处理, 则每一路卷积神经网络的特征向量个数不宜过多, 一方面不利于特征联合, 另一方面会造成特征信息冗余, 反而影响分类效果. 所以我们对第 7 层及之后的全连接层的层数和输出进行了探索 (详见第 3 节表 2), 最终设计了一个 11 层的深度卷积神经网络, 如图 2 所示.

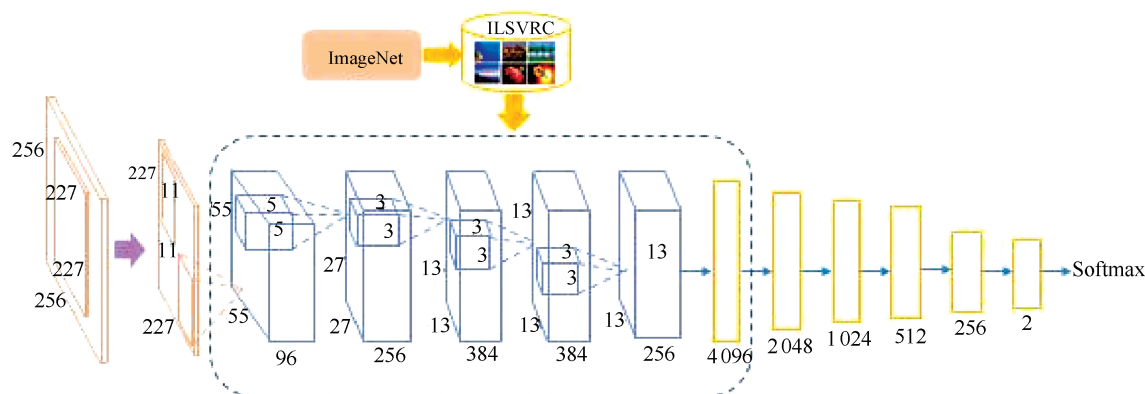


图 2 单路卷积神经网络

Fig. 2 Single column convolutional neural networks

第 7 层全连接层, 有 2048 个输出; 第 8 层全连接层, 有 1024 个输出; 第 9 层全连接层, 有 512 个输出; 第 10 层全连接层, 有 256 个输出; 第 11 层全连接层的输出个数和类别输出个数相同, 均为 2. 考虑到网络层数越深, 其特征信息的代表性越强, 因此我们将提取第 10 层全连接层的特征作为图像的特征向量.

该网络结构的输入图像大小为 256 像素 $\times$ 256 像素. 这两个数值分别是宽度和高度. 图像数据被随机裁剪出 10 个 227 像素 $\times$ 227 像素大小的图像, 保证这些输入能尽量覆盖到一幅图像的全部区域.

在网络参数初始化阶段, 我们的做法为: 利用文献 [13] 的深度网络训练图像分类数据库得到的网络模型的前 5 层卷积层和第 6 层全连接层的网络参数, 参与本文深度卷积神经网络中对应网络层的参数初始化. 而本文网络的第 7 层至第 11 层网络层的参数则采用随机初始化. 这是因为卷积层特征代表的是图像的特征信息, 随着卷积层层数的加深, 特征信息不断抽象. 而全连接层可以理解为是一维列向量对卷积层特征图的总结变换, 不同类别图像区别较大, 针对性强. 总的来说, 网络层数越深, 其参数与既定分类任务越相关; 而网络层数越浅, 其参数对特征表达来说越具体. 所以, 我们利用文献 [13] 的深度网络训练图像分类数据库得到的网络模型参与本文网络前 6 层的参数初始化.

训练网络时, 网络中的参数随输出和样本标签计算所得的损失函数不断更新. 我们将前 6 个网络层的学习率设为 0.001, 后面的网络层的学习率设为 0.005, 以使随机初始化的全连接层参数更新比其他层快. 学习率随着迭代次数的增加都会有规律地减少, 具体设置与训练数据集的样本数和训练中的梯度下降算法有关.

2.2 并行网络的输入

影响图像美感的因素是多方面的, 构图、色彩、亮度、纹理、复杂度等对图像美感都有影响. 对同一幅图像, 不同因素对其美感质量的影响程度是不同的. 如图 3 所示, 色彩是影响图 3(a) 美感的主要因素, 构图是影响图 3(b) 美感的主要因素, 亮度是影响图 3(c) 美感的主要因素, 纹理是影响图 3(d) 美感的主要因素. 图像的编码方式有多种, 每种方式都侧重图像的不同方面的描述. 将同一幅图像不同的表达形式输入到深度学习网络中, 网络学习到的特征信息不同. 将这些特征联合起来, 可以更全面地表达图像信息. 于是, 我们从图像的构图、色彩、饱和度、亮度、纹理细节等方面考虑并行卷积神经网络的输入, 具体做法如下:

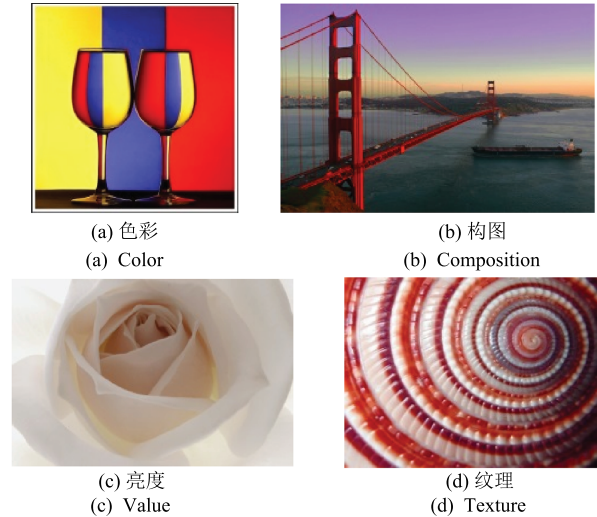


图 3 影响图像美感的主要因素示例

Fig. 3 The main factors affecting aesthetics of images

1) Normal: RGB 图像从不同颜色分量混合的角度以色彩来表示图像. 由于网络要求输入是宽高相同的图像, 因此我们把 RGB 图像缩放到 256 像素 $\times$ 256 像素的大小作为网络的输入.

2) Resize: 若图像本来的宽高不同, 直接对图像进行缩放, 会改变图像原来的尺寸比例, 导致图像产生畸变, 这对图像美感会造成一定的影响. 我们采取以下做法: 把图像的长边固定变换到 256, 短边则按长边的变换比例进行变换, 图像中空余出来的位置则用 0 填充.

3) H、S、V: HSV 图像的 H、S 和 V 三个通道分别代表图像的色度、饱和度和亮度信息, 和 RGB 图像的三通道代表的意义区别较大. 我们将 RGB 图像转换为 HSV 图像, 并把 HSV 图像的单个通道矩阵取出, 即 H 通道图像矩阵、S 通道图像矩阵、V 通道图像矩阵, 分别作为深度学习网络的输入.

4) Daubechies: 图像能量中含有丰富的边缘、轮廓信息, 图像的小波系数包含丰富的能量信息. 已有图像美感分类相关的论文通过图像的 Daubechies 小波变换, 对小波系数进行建模计算得到图像高层美学特征^[3,6]. 本文中, 我们将图像进行 Daubechies 小波变换, 取其第 1 层小波系数矩阵作为我们深度学习网络的输入.

各种图像输入示例如图 4. 为了验证并确定哪些输入以及几层并行组合是有效的, 我们利用这些不同形式的图像矩阵作为单路卷积神经网络的输入, 组建了并行卷积神经网络. 分别进行了多种不同输入的网络并行实验 (详见第 3.2 节), 得到了多种输入并行下的分类结果. 最后从算法的分类准确率和算法复杂度方面综合考虑, 选择 Normal 和 V 共同组建了双路并行卷积神经网络.



图4 卷积神经网络的不同图像输入形式示例

Fig.4 Examples of different types of input images of convolutional neural networks

2.3 特征结合与图像分类

我们分别将以上图像矩阵输入到深度卷积神经网络中, 进行训练得到相应的网络模型. 对每一幅图像分别利用以上模型提取第 10 层 (全连接层) 的特征向量, 将这些特征向量进行组合, 得到完整的图像美学特征.

由于不同的特征组合会导致特征数量不同, 也考虑到在各美感图库中特征和图像数量变化较大, 我们使用泛化性能较好的支持向量机 (SVM) 机器学习方法对这些组合特征进行学习, 得到最终的图

像美感分类器.

3 实验与分析

为了评估本文方法的有效性, 我们分别在 AVA^[20] 和 CUHKPQ^[9] 两个不同的图像美感数据库上进行了实验, 本节给出实验结果与分析.

3.1 图像美感分类数据库

AVA 数据库和 CUHKPQ 数据库是在图像美感分类中的常用数据库, 其中的图像都是来自受欢迎的社交网站中人们共享的图像. 图 5 是 AVA 数据库和 CUHKPQ 数据库中的高美感和低美感图像示例.

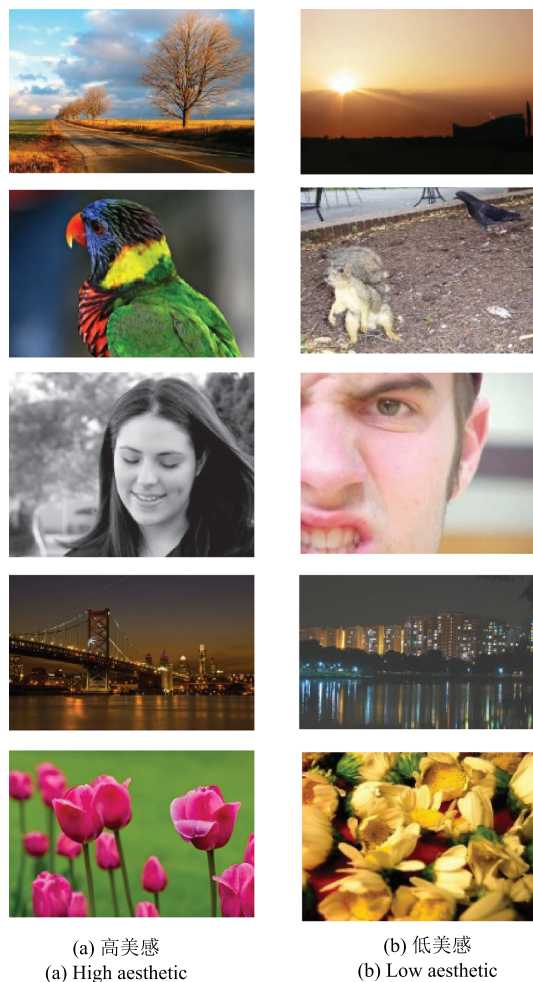


图5 数据库中高美感和低美感图像示例

Fig.5 Examples of high aesthetic images and low aesthetic images of datasets

3.1.1 AVA 数据库

AVA^[20] 是一个包含超过 25 万幅来自社交图片分享网站 www.dpchallenge.com 的图像的大规模数据库. 每幅图像采集了平均 210 个用户对图像的

主观美感评分, 评分范围是 1~10 分. 这些评分者包含各种群体, 不分性别、年龄和专业限制, 保证了评分的客观性和广泛代表性. 每幅图像各用户评分的平均值作为图像的美感评分标签, 分数越高, 代表图像的美感越高. 文献 [20] 提供了 AVA 数据库中图像的网络链接供我们下载, 据此我们共收集了 255 529 幅图像构成本文的 AVA 数据库.

我们对图像的美感评分进行排序, 利用以下两种方式对数据库进行高低美感类别划分.

1) AVA1. 为了减少图像样本噪声和增加高低美感样本的类间距离, 我们按照 Dong 等^[16] 的做法, 挑选图像数据库中前 10% 高分的图像作为高美感图像, 并挑选前 10% 低分的图像作为低美感图像, 高低美感图像分别有 25 532 幅. 然后随机选取高低美感类图像中的一半作为训练集, 另一半作为测试集. 下文我们以 AVA1 表示此图库.

2) AVA2. 按照 AVA 数据库^[20] 给出的训练和测试数据集的划分, 得到 235 599 幅图像作为训练数据集, 19 930 幅图像作为测试数据集. 按照 AVA 数据库^[20] 和 Lu 等^[10] 的做法, 以 5 分为界, 将训练集和测试集中的图像划分为高美感和低美感两个类别. 美学评分为 5 分以上的归类为高美感图像, 美学评分为 5 分以下的归类为低美感图像. 如此, 我们得到 180 856 幅高美感图像和 74 673 幅低美感图像. 其中, 训练集包含高美感图像 166 689 幅, 低美感图像 68 910 幅; 测试集包含高美感图像 14 167 幅, 低美感图像 5 763 幅. 下文我们以 AVA2 表示此图库.

3.1.2 CUHKPQ 数据库

CUHKPQ 是一个包含 17 673 幅来自专业摄影网站图像的数据库^[9]. 且根据图像场景分成了 7 个类别, 分别为 Animal、Plant、Static、Architecture、Landscape、Human 和 Night. 在数据库构建时, 每幅网络图像由 10 个用户进行高低美感的二分类评价, 有 8 个或以上的用户给出评价相同的图像才选入数据库. 这有效地减少了图像库的噪声.

本文在 CUHKPQ 各场景类别图库和包含各类别所有图像的总图库 (Overall) 上进行实验. 为了方便图像美感分类方法的比较, 本文参照文献 [9, 12, 16] 的方法, 将各类别图库中一半的高美感图像和一半的低美感图像作为训练集, 另一半作为测试集.

3.2 并行深度卷积神经网络结构实验

3.2.1 单路卷积神经网络结构

为了更好地适应本文并行卷积神经网络特征的合并, 即减少单路网络的特征向量的维度, 以减少特征的冗余性和算法的复杂度, 同时又要保证单路卷积神经网络模型对输入图库的分类性能, 本节主要探讨本文单路卷积神经网络的结构设计.

网络结构的 5 层卷积层是不变的, 与文献 [13] 网络结构的卷积层同. 我们主要对全连接层的层数和输出特征值的个数进行了设计, 并在 AVA1 数据集上进行了实验, 结果如表 1 所示.

表 1 不同结构单路卷积神经网络的分类准确率
Table 1 Classification accuracy of single column convolutional neural networks with different structures

全连接 层设置	Fc 4 096	Fc 2 048	Fc 1 024	Fc 512	Fc 256	Fc 2	分类准确率 (%)
Arch1	√	√				√	83.70
Arch2	√	√	√			√	83.73
Arch3	√	√	√	√		√	83.21
Arch4	√	√	√	√	√	√	83.28

表 1 中, 每种网络结构包含的层用 “√” 表示, Fc 表示全连接层, 下面的数值表示该全连接层的输出特征值个数.

表 1 显示, 全连接层层数的增加对美感分类效果的影响差别不大. Arch2 只有 3 层的结构可以获得最高的准确率 83.73%, 但得到的特征维数太高 (1 024 维), 不利于后期处理. 综合考虑准确率和特征维数, 本文选择 Arch4 作为我们的单路卷积神经网络结构. 在 Arch4 中, 第 10 层的输出特征值个数为 256, 特征维数适当, 有利于特征合并.

3.2.2 并行网络结构

本节主要探索最有效的并行网络结构, 即将第 2.2 节中哪些图像矩阵输入到并行网络中并进行组合, 能更好地学习图像美感信息, 得到更好的美感分类效果.

首先对各种单路网络输入进行实验. 我们利用 AVA1 数据集, 分别将第 2.2 节提出的各种图像矩阵输入到单路深度卷积神经网络中, 进行图像美感分类的训练和测试, 得到了以下分类准确率, 如表 2 所示.

可以看出: 1) 将图像直接缩放到宽高分别为 256 像素的图像的处理, 美感分类的准确率最高, 为 83.28%; 2) 取 HSV 图像的一个通道 V 的矩阵, 能达到 82.99% 的准确率; 3) Daubechies 小波系数矩阵能达到 81.60% 的准确率. 可见, 图像的颜色、能量、亮度都是影响图像美感的重要因素.

接着, 我们确定哪些输入组合效果最好. 我们利用各个单路深度卷积神经网络训练出来的模型, 提取图像的特征, 将各路特征进行组合, 利用 SVM 进行训练和测试. 各种特征组合方式的分类准确率如表 3 所示, 其中, 每种组合方式所包含的特征类型用

“√”表示.

表 2 不同输入的单路卷积神经网络的分类准确率
Table 2 Classification accuracy of single column convolutional neural networks with different inputs

输入方式	分类准确率 (%)
Normal	83.28
Resize	80.28
H	70.03
S	75.90
V	82.99
Daubechies	81.60

表 3 各种特征组合方式的分类准确率
Table 3 Classification accuracy of various features combination

输入组合	Normal	Resize	H	S	V	Daubechies	特征维数	分类准确率 (%)
1			√	√	√		768	83.93
2	√						256	83.28
3	√		√				512	83.66
4	√			√			512	84.18
5	√				√		512	85.00
6	√		√	√			768	85.17
7	√			√	√		768	85.33
8	√			√		√	768	85.83
9	√	√		√	√		1 024	85.41
10	√		√	√	√	√	1 280	85.94

由以上结果可以看出: 1) 将特征组合的图像美感分类准确率都比原单路特征的分类准确率有所提升. 2) 将从 Normal 和 V 输入矩阵所提取的特征组合能达到 85.00% 的准确率 (第 5 种). 3) 从表 3 中可以看出, 第 10 种特征组合方式所得的美感分类准确率最高, 为 85.94%. 但其特征维数也较高, 为 1 280 维. 这导致了算法复杂度的提高, 与第 5 种特征组合方式相比, 特征维数增加了一倍以上, 而准确率没有很大的提高. 4) 第 6、7、8 种特征组合方式, 分别为 Normal、H、V 进行组合; Normal、S、V 进行组合; Normal、V、Daubechies 进行组合, 特征维数都为 768, 比第 5 种特征组合方式稍高, 图像美感分类准确率也比第 5 种特征组合方式稍高. 但 Daubechies 图像矩阵的算法复杂度较其他图像矩阵大.

在样本数量有限的数据库中, 特征维数过高, 不利于模式识别问题的解决. 在有时间效率要求的情况下, 算法复杂度会影响算法的实际应用. 本文综合考虑图像美感分类准确率、算法复杂度和对图像美感数据库的适用性问题, 最终选定第 5 种特征组合方案, 即将 Normal 和 V 作为输入的两路深度卷积神经网络所提取的图像美感特征进行组合. 在实际图像美感分类应用的过程中, 如果时间效率允许, 且图像美感训练数据充足, 可以考虑采用第 6、7、8 种特征组合方案.

4 图像美感分类实验结果与对比

本节中我们展示本文方法在上文所述图像美感分类数据库上的效果, 并与传统手工特征、局部特征提取方法和现有深度学习图像美感分类方法的分类效果进行比较.

4.1 AVA 数据库实验结果与对比

4.1.1 AVA1 数据库

在 AVA1 数据库中, 为了方便图像美感分类算法的对比, 训练数据集和测试数据集的设置与文献 [16] 相同. 我们利用本文算法, 用训练集中高、低美感图像各 12 766 幅训练得到美感分类器, 并对测试数据集中高、低美感的图像各 12 766 幅进行图像美感分类实验. 与现有方法的对比结果如表 4 所示.

表 4 AVA1 数据库的实验结果及与现有方法的对比
Table 4 The experimental results of the AVA1 datasets and comparison with existing methods

方法类型	图像美感分类方法	分类准确率 (%)
手工特征	Datta 等 ^[6]	68.67*
	Ke 等 ^[8]	71.06*
局部特征	Marchesotti 等 ^[11]	68.55*
深度学习方法	DCNN_Aesth_SP ^[16]	83.52
	本文方法	85.00

* 此行数据引用自文献 [16] 的结果.

从表 4 可以看到, 本文方法的图像美感分类准确率最高, 为 85%. 即高于手工特征的方法^[6,8], 也高于 Marchesotti 等^[11] 的局部特征的方法. 即使是最新的深度学习方法的运用, 也比 Dong 等^[16] 的方法的准确率高约 1.5%. 总体来说, 深度学习的方法要比传统特征提取的方法性能更好.

4.1.2 AVA2 数据库

在利用深度学习方法进行图像美学分类的课题中, Lu 等^[10] 也进行了研究. 因此, 我们在 AVA2 数

据库上利用本文的图像美感分类算法进行实验, 并与 Lu 等^[10] 的方法进行了对比, 结果如表 5 所示.

表 5 AVA2 数据库的实验结果及与现有方法的对比
Table 5 The experimental results of the AVA2 datasets and comparison with existing methods

图像美感分类方法	分类准确率 (%)
RDCNN_semantic ^[10]	75.42
本文方法	77.03

由表 5 中的实验数据可以看出, 本文方法的分类准确率比 Lu 等^[10] 方法的高.

总的来说, 深度学习方法要比传统手工特征方法的图像美感分类性能更好. 而本文方法在两种 AVA 数据库的实验中, 图像美感分类的效果最好.

4.2 CUHKPQ 数据库实验结果与对比

对 CUHKPQ 图像美感数据库, 我们分别在 7 个不同场景类别以及总图库上利用本文算法进行了实验. 先利用数据集对深度卷积神经网络模型进行训练, 确定网络模型参数后, 提取图像特征并组合, 再利用 SVM 算法训练图像美感分类器, 对测试数据集图像进行图像美感分类得到最终的分类准确率.

为了与其他的图像美感分类方法进行对比, 本文按照文献 [9]、文献 [12] 和文献 [16] 的做法, 随机选取数据库中一半的高美感图像和低美感图像作为训练数据集, 另一半图像作为测试数据集. 为保证实验结果的客观性, 我们重复 10 次实验, 取 10 次实验图像美感分类准确率的平均值作为最终的分类准确率.

图像美感分类准确率以及与传统手工特征、局部特征及现有深度学习方法对比的结果, 如表 6 所示.

表 6 CUHKPQ 各类别图库和总图库的实验结果及现有方法的对比
Table 6 The experimental results of the CUHKPQ datasets and comparison with existing methods

特征类型		场景类别							Overall
		Animal	Architecture	Human	Landscape	Night	Plant	Static	
手工特征	All features in [8]*	0.7751	0.8526	0.7908	0.8170	0.7321	0.8093	0.7829	0.7944
	All features in [9]	0.8937	0.9275	0.9740	0.9468	0.8463	0.9182	0.9069	0.9209
局部特征	Semantic features ^[12]	0.8623	0.8644	0.9313	0.8416	0.8742	0.8685	0.8964	0.8787
	Semantic features + handcrafted features ^[12]	0.9033	0.8755	0.9472	0.8853	0.9052	0.9232	0.9094	0.9093
深度学习方法	DCNN_Aesth_SP ^[16]	—	—	—	—	—	—	—	0.9193
	本文方法	0.9382	0.9113	0.9697	0.9100	0.9166	0.9410	0.9159	0.9395

* 此行数据引用自文献 [12] 的结果.

示.

在表 6 中, 本文方法与手工特征提取方法^[8-9]、局部特征提取方法^[12] 以及已有的深度学习网络方法^[16], 进行了比较.

在总图库 (Overall) 上, 本文方法的图像美感分类准确率最高, 达到了 93.95%. 与手动特征方法比较, 本文方法明显高于早期手动特征方法^[8], 比 Tang 等^[9] 方法的准确率高近 2%. 与局部特征方法的比较上, 比 Guo 等^[12] 方法的准确率高约 3%. 在与 Dong 等^[16] 的最新深度学习方法的比较上, 效果也更好, 高约 2%.

在 CUHKPQ 各场景类别图库的图像美感分类准确率上, 除了在 3 个类别 (Architecture、Human、Landscape) 的准确率上稍稍低于 Tang 等^[9] 的结果之外, 本文方法均高于其他方法. 由于深度学习方法一般采用较深层次的网络结构对数据进行分析, 则对样本数据的数据量有一定要求, 深度学习方法擅长处理大量样本数据的分析问题. CUHKPQ 图库的各单类别图库样本数量偏少, 且不同类别的样本数量存在不平衡的问题, 深度学习方法不能很好地发挥优势.

综上所述, 在多个数据库上的图像美感分类实验结果表明, 本文提出的并行卷积神经网络的图像美感分类方法, 比传统的手工特征、局部特征和现有的深度学习方法有更好的美感分类效果.

5 总结与展望

本文从影响图像美感的不同角度出发, 对图像的描述矩阵、深度卷积神经网络的结构和并行深度卷积神经网络进行了研究, 提出一种双路并行深度卷积神经网络, 应用于图像美感分类问题.

深度学习网络解决了传统手工特征和局部特征难以对图像美学特征量化建模的问题, 可以直接从图像中学习得到美学特征. 本文利用深度学习网络的特征学习能力, 综合考虑影响图像美感的颜色、亮度、构图等信息, 设计了并行深度卷积神经网络, 借用了大规模图像分类数据集训练所得的模型中所包含的图像信息, 设计了不同的图像描述矩阵作为网络的输入, 最后将特征有效组合. 主要贡献如下: 1) 利用大规模图像分类和识别数据库 ImageNet 训练得到的模型^[13], 参与本文深度学习网络模型的权值初始化, 以解决当图像美感数据集中的样本数量不足时, 深度学习网络的特征学习能力受到限制的问题; 2) 对图像设计了不同的网络输入表达矩阵, 以从影响图像美感的不同方面考虑, 更全面地学习图像美感特征. 从实验结果分析与对比来看, 本文算法取得了比图像美感评价中现有的传统手工特征、局部特征或深度学习方法更好的美感分类效果.

当然, 并行深度卷积神经网络用于图像美感评价研究也还存在一些可以思考 and 发展的地方, 例如: 更好的图像表达形式、更合理有效的特征组合方式、在深度学习网络中利用显著区域^[21]等. 又如怎样能使深度学习网络学习到不同场景下图像的美感信息, 让网络有更广泛的针对性和适用性, 这些都是我们接下来的研究发展方向.

References

- 1 Wang Wei-Ning, Yi Jing-Jian, He Qian-Hua. Review for computational image aesthetics. *Journal of Image and Graphics*, 2012, **17**(8): 893–901
(王伟凝, 蚁静斌, 贺前华. 可计算图像美学研究进展. 中国图象图形学报, 2012, **17**(8): 893–901)
- 2 Wang Wei-Ning, Liu Jian-Cong, Xu Xiang-Min, Jiang Yi-Zi, Wang Li. Aesthetic enhancement of images based on photography composition guidelines. *Journal of South China University of Technology (Natural Science Edition)*, 2015, **43**(5): 51–58
(王伟凝, 刘剑聪, 徐向民, 姜怡孜, 王励. 基于构图规则的图像美学优化. 华南理工大学学报 (自然科学版), 2015, **43**(5): 51–58)
- 3 Wang Wei-Ning, Yi Jing-Jian, Xu Xiang-Min, Wang Li. Computational aesthetics of image classification and evaluation. *Journal of Computer-Aided Design & Computer Graphics*, 2014, **26**(7): 1075–1083
(王伟凝, 蚁静斌, 徐向民, 王励. 可计算的图像美学分类与评估. 计算机辅助设计与图形学学报, 2014, **26**(7): 1075–1083)
- 4 Wang W N, Cai D, Wang L, Huang Q H, Xu X M, Li X L. Synthesized computational aesthetic evaluation of photos. *Neurocomputing*, 2016, **172**: 244–252
- 5 Tong H H, Li M J, Zhang H J, He J R, Zhang C S. Classification of digital photos taken by photographers or home users. In: *Proceedings of the 5th Pacific Rim Conference on Multimedia*. Tokyo, Japan: Springer, 2004. 198–205
- 6 Datta R, Joshi D, Li J, Wang J Z. Studying aesthetics in photographic images using a computational approach. In: *Proceedings of the 9th European Conference on Computer Vision*. Graz, Austria: Springer, 2006. 288–301
- 7 Wang W N, Zhao W J, Cai C J, Huang J X, Xu X M, Li L. An efficient image aesthetic analysis system using Hadoop. *Signal Processing: Image Communication*, 2015, **39**: 499–508
- 8 Ke Y, Tang X O, Jing F. The design of high-level features for photo quality assessment. In: *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. New York, USA: IEEE, 2006. 419–426
- 9 Tang X O, Luo W, Wang X G. Content-based photo quality assessment. *IEEE Transactions on Multimedia*, 2013, **15**(8): 1930–1943
- 10 Lu X, Lin Z, Jin H L, Yang J C, Wang J Z. Rating image aesthetics using deep learning. *IEEE Transactions on Multimedia*, 2015, **17**(11): 2021–2034
- 11 Marchesotti L, Perronnin F, Larlus D, Csorika G. Assessing the aesthetic quality of photographs using generic image descriptors. In: *Proceedings of the 2011 IEEE International Conference on Computer Vision*. Barcelona, Spain: IEEE, 2011. 1784–1791
- 12 Guo L H, Xiong Y C, Huang Q H, Li X L. Image esthetic assessment using both hand-crafting and semantic features. *Neurocomputing*, 2014, **143**: 14–26
- 13 Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks. In: *Proceedings of the 26th Annual Conference on Neural Information Processing Systems 2012*. Lake Tahoe, USA: Curran Associates, Inc., 2012. 1097–1105
- 14 Sun Y, Wang X G, Tang X O. Deep learning face representation from predicting 10 000 classes. In: *Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition*. Columbus, USA: IEEE, 2014. 1891–1898
- 15 Lee H, Grosse R, Ranganath R, Ng A Y. Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *Proceedings of the 26th Annual International Conference on Machine Learning*. Montreal, Canada: ACM, 2009. 609–616
- 16 Dong Z, Shen X, Li H Q, Tian X M. Photo quality assessment with DCNN that understands image well. In: *Proceedings of the 21st International Conference on MultiMedia Modeling*. Sydney, Australia: Springer International Publishing, 2015. 524–535
- 17 Dong Z, Tian X M. Multi-level photo quality assessment with multi-view features. *Neurocomputing*, 2015, **168**: 308–319
- 18 Deng J, Dong W, Socher R, Li L J, Li K, Li F F. ImageNet: a large-scale hierarchical image database. In: *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, USA: IEEE, 2009. 248–255
- 19 Yin W, Mei T, Chen C W. Assessing photo quality with geo-context and crowdsourced photos. In: *Proceedings of the 2012 IEEE Visual Communications and Image Processing*. San Diego, USA: IEEE, 2012. 1–6
- 20 Murray N, Marchesotti L, Perronnin F. AVA: a large-scale database for aesthetic visual analysis. In: *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*. Providence, USA: IEEE, 2012. 2408–2415

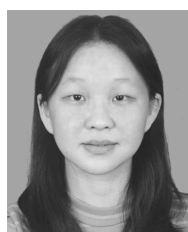
- 21 Wang W N, Cai D, Xu X M, Liew A W C. Visual saliency detection based on region descriptors and prior knowledge. *Signal Processing: Image Communication*, 2014, **29**(3): 424–433



王伟凝 华南理工大学电子与信息学院副教授. 2005 年获得华南理工大学通信与信息系统专业博士学位. 主要研究方向为图像处理与模式识别, 计算机视觉, 机器学习.

E-mail: wnwang@scut.edu.cn

(**WANG Wei-Ning** Associate professor at the School of Electronic and Information Engineering, South China University of Technology. She received her Ph.D. degree from South China University of Technology in 2005. Her research interest covers image processing, pattern recognition, computer vision, and machine learning.)



王 励 华南理工大学电子与信息学院硕士研究生. 主要研究方向为图像处理, 机器学习与计算机视觉.

E-mail: li_wang72@163.com

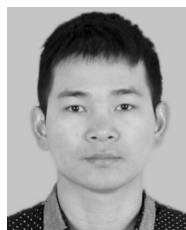
(**WANG Li** Master student at the School of Electronic and Information Engineering, South China University of Technology. Her research interest covers image processing, machine learning, and computer vision.)



赵明权 华南理工大学电子与信息学院硕士研究生. 主要研究方向为计算机视觉, 机器学习.

E-mail: yomzhao@gmail.com

(**ZHAO Ming-Quan** Master student at the School of Electronic and Information Engineering, South China University of Technology. His research interest covers computer vision and machine learning.)



蔡成加 华南理工大学电子与信息学院硕士研究生. 主要研究方向为图像处理, 机器学习和计算机视觉.

E-mail: caichgjia@163.com

(**CAI Cheng-Jia** Master student at the School of Electronic and Information Engineering, South China University of Technology. His research interest covers image processing, machine learning, and computer vision.)



师婷婷 华南理工大学电子与信息学院硕士研究生. 2015 年获郑州大学电子信息科学与技术学士学位. 主要研究方向为图像处理, 计算机视觉与模式识别.

E-mail: eestt.ll@mail.scut.edu.cn

(**SHI Ting-Ting** Master student at the School of Electronic and Information Engineering, South China University of Technology. She received her bachelor degree from Zhengzhou University in 2015. Her research interest covers image processing, computer vision, and pattern recognition.)



徐向民 华南理工大学电子与信息学院教授. 2001 年获华南理工大学电子与信息学院博士学位. 主要研究方向为图像/视频处理, 人机交互, 计算机视觉与机器学习. 本文通信作者.

E-mail: xmxu@scut.edu.cn

(**XU Xiang-Min** Professor at the School of Electronic and Information Engineering, South China University of Technology. He received his Ph.D. degree from the School of Electronic and Information Engineering, South China University of Technology in 2001. His research interest covers image/video processing, human-computer interaction, computer vision, and machine learning. Corresponding author of this paper.)