

基于深度学习的照片图像美感质量评估综述

李文轩

(浙江理工大学信息学院, 浙江 杭州 310018)

✉xuchange2022@163.com



摘 要: 图像美感质量评估作为计算美学中重要的研究领域,是指利用计算机模拟人类的主观审美思维,并在此基础上对图像的美感进行定性或定量评估。作为图像美感质量评估中的一类主要研究对象,照片图像美感质量评估在检索与排序、照片图像美化等方面有着广泛的应用。本文主要对基于深度学习技术的照片图像美感质量评估研究进行归纳和总结,并从基于深度学习的美感评估这一基本思路出发,依次对照片图像美感质量评估类别、照片图像美感评估数据集及其建立方法进行综述,并对相关研究内容提出建议。

关键词: 计算美学;深度学习;照片图像美感质量评估;评估类别;数据集

中图分类号: TP391 **文献标识码:** A

Overview of Aesthetic Quality Evaluation of Photo Images based on Deep Learning

LI Wenxuan

(School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China)

✉xuchange2022@163.com

Abstract: As an important research field in computational aesthetics, image aesthetic quality evaluation refers to the use of computer to simulate human subjective aesthetic thinking, and on this basis, to carry out a qualitative or quantitative evaluation of the image aesthetics. Photo image aesthetic quality evaluation, as one of the main research objects in image aesthetic quality evaluation, has been widely applied in retrieval and sorting, photo image beautification and so on. This paper mainly proposes to summarize the research on the aesthetic quality evaluation of photo images based on deep learning technology. Starting from the basic idea of the aesthetic evaluation based on deep learning, categories of photo image aesthetic quality evaluation, photo image aesthetic evaluation datasets and establishment methods are reviewed in turn, and suggestions for relevant research are put forward.

Keywords: computational aesthetics; deep learning; photo image aesthetic quality evaluation; evaluation categories; dataset

1 引言(Introduction)

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,以及美的创造、发展及其规律的科学^[1]。而使用人工智能技术对图像进行定量或定性的评估则被称为“计算美学^[2]”,即通过研究“美”的可计算方法,使得计算机能够对图像做出符合人类倾向的审美决策,该概念在

2005年欧洲计算机图形学会(EG)举办的第一届图形图像视觉计算美学学术会议上被正式提出。

照片图像美感质量评估是“计算美学”研究的重点领域。随着照相机、智能手机等摄影成像设备的快速普及,在线照片图像的数据量呈指数式增长,产生了繁杂且冗余的海量在线照片图像,因此,如何在海量照片图像中获取更高质

量美感的照片图像显得尤为重要。使用人工智能技术模拟人的审美感知,并对海量照片图像高效地进行美感质量评估可以有效地解决该类问题。

2 基于深度学习的照片图像美感质量评估基本思路(Basic ideas of photo image aesthetic quality evaluation based on deep learning)

本文总结了以照片图像为美感评估对象,以美感质量为美感评估任务,以自动特征提取方法为主的基于深度学习的照片图像美感质量评估基本思路,如图1所示。

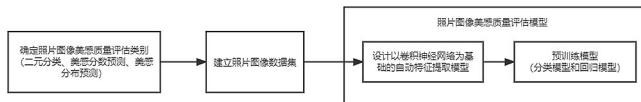


图1 基于深度学习的照片图像美感质量评估基本思路

Fig.1 Basic ideas of photo image aesthetic quality evaluation based on deep learning

实际上,在研究过程中,往往对自动特征提取和预测模型进行统一的设计和训练,实现“端到端”的学习模式。并且,在确立照片图像美感质量评估类别阶段,往往也同时考虑照片图像美感质量评估模型,因此,本文不再单独对照片图像美感质量评估模型进行综述。

3 基于深度学习的照片图像美感质量评估类别(Evaluation categories of photo image aesthetic quality based on deep learning)

基于深度学习的照片图像美感质量评估类别可以分成三类:第一类是二元分类,即把照片图像美感质量按照高美感和低美感进行定性划分;第二类是美感分数预测,即以连续数值的形式对照片图像美感质量进行量化;第三类是美感分布预测,即在美感分数预测的基础上预测照片图像美感分数的概率分布,得到相应的直方图,并根据概率论和数理统计的知识计算出平均值、方差等数据。

3.1 二元分类

二元分类是早期照片图像美感质量评估研究的主要内容。2014年,随着卷积神经网络在图像分类问题上被广泛使用并表现出优异的性能,LU等人^[3]首次尝试将卷积神经网络应用于二元分类的图像美感质量评估研究,开创了自动提取特征方法的先河,并在进行模型训练时使用随机裁剪方法加入图像的局部信息。

TIAN^[4]通过加入跳连操作,提出了一种基于深度和浅层特征融合的图像美感评估模型。但是以上方法的模型输入都是固定 256×256 大小的照片图像,因此针对不同大小的照片图像需要进行裁剪,会在一定程度上降低照片图像的分辨率甚至导致照片图像失真,影响美感质量评估的模型性能。MA等人^[5]提出一种自适应感知的卷积神经网络,通过设计算法提取多个图像的局部特征块作为神经网络的输入,从而使得神经网络模型接受任意大小的图像输入。他们尽管实现了任意图像大小的输入,但是通过局部图像块的方式又会使图像的全局结构信息丢失。实际上,输入图像大小固定的原因在于输入最后一层全连接的特征维度必须固定,MAI等人^[6]在此基

础上通过增加自适应空间池层固定最后的输出维度,使得输入图像的大小可以不受限制。不难发现,为了提高二元分类模型的性能,早期研究者主要围绕局部图像块与全局结构信息之间的矛盾进行研究。

3.2 美感分数预测

随着基于深度学习的照片图像美感质量评估研究的深入,二元分类的缺陷也愈发明显,尽管其可以区分高美感图片和低美感图片,但是无法应用于照片图像质量排序、照片图像增强等场景。并且,相较于二元分类,回归模型预测美感分数也更接近人类视觉系统中美感评估的过程^[7]。2015年,KAO等人^[8]首次将卷积神经网络应用于美感分数预测的图像美感质量评估研究。但是由于图像美感分数的评分数据相较于二值分类更难获得,建立数据集的代价比较昂贵,为了减小建立数据集方面的工作量,HUANG等人^[9]使用迁移学习的方法将分类任务学习到的特征迁移到回归模型中,该方法在减小工作量的同时提高了模型的性能。但是实际上,考虑到美感评分的主观差异,现今的工作大都偏向于直接预测美感分布,通过美感分布间接算出美感分数。

3.3 美感分布预测

美感分布预测是对美感分数预测的进一步深入,即预测人类审美评估间的概率分布,并最终直方图的形式进行呈现,同时利用概率论与数理统计的知识,在此基础上计算出均值、方差、中位数、偏度、峰度等一系列数据,最终从各个角度对照片图像美感进行评估。其考虑到了人类审美评估的主观性,避免了用一个连续量表示照片图像美感所带来的弊端。2018年,TALEBI等人^[10]首次将卷积神经网络应用于美感分布预测。他们使用迁移学习的方法,选择性能优异的图像分类卷积神经网络模型,将其最后一层分类层修改为包含十个类别的全连接层,十个类别对应十个美感分数区间,最后经Softmax输出美感分布。

TALEBI等人还同时考虑到了美感分布预测的类间关系,例如第一个美感分数区间与第二个美感分数区间之间的联系会比第一个美感分数区间与第三个美感分数区间之间的联系更紧密,因此提出使用EMD分布距离损失函数取代交叉熵损失函数对模型进行训练,最终的结果表明,EMD损失函数在美感分布预测任务中表现出的性能优于交叉熵损失函数。最后,TALEBI等人通过实验结果证明,使用美感分布预测的结果计算平均值,从而间接得到的美感分数相较于直接预测美感分数的准确率更高,但是根据该方法计算出来的方差准确性欠佳。CUI等人^[11]提出了一种语义感知混合网络,通过收集图像中的内容信息,比如图像分类和场景类别,提高美感分布预测的性能。WANG等人^[12]提出了一种能够保留原始图像纵横比信息的美感分布预测方法,与CUI等人的方法相比,在没有加入额外信息(比如图像内容信息)的同时,仍保持较好的性能,因此该方法对其他学习任务有普适性意义。不难发现,自TALEBI等人将深度学习应用于美感分布预测之后,该研究方向的重点逐渐集中于提出不同的神经网络结构,加入更多的图像信息,提高模型本身的性能。

4 照片图像美感质量评估数据集(Photo image aesthetic quality evaluation datasets)

4.1 AVA数据集

MURRAY等人^[13]在2012年建立了照片图像美感评估领域第一个大型数据集——AVA照片图像美感评估数据集。AVA数据集收集了大约255,000张照片图像,每张照片图像分别有三类注释:第一类是美感评分,每张照片图像由78—549人进行评分,平均每张照片图像有210名评分者,评分范围为1—10分,并最终根据评分的平均值得到照片图像的美感评分数据,评分者的身份包括专业摄影师和业余爱好者。第二类是语义标签,具体为家庭、建筑、天空等66个语义描述。第三类是照片图像属性,研究者根据光线、颜色、构图三种关键要素设计和选择互补色、双色等14个属性对照片图像进行描述。

AVA数据集来源于网站www.dpchallenge.com。该网站有一种“挑战”机制,即给出一个挑战标题和简短的描述,参与者上传符合挑战要求的图片并接受全网站用户的评分,最终根据每张图片的平均得分对其进行排序,如图2所示。

Challenge Results for Through the Looking Glass		
	Legs And Leaves by DofflerDDS	Your Vote: N/A Your Comment: N/A Average Vote (Participants): 6.889 Average Vote (Non-Participants): 7.000
1st place with an average vote of 6.9667		
Views: 135 (69 during voting) Votes: 30		
	morning by Tiberius	Your Vote: N/A Your Comment: N/A Average Vote (Participants): 7.250 Average Vote (Non-Participants): 6.591
2nd place with an average vote of 6.7667		
Views: 58 (76 during voting) Votes: 30		
	Tribute to Erwin by Lev7	Your Vote: N/A Your Comment: N/A Average Vote (Participants): 7.000 Average Vote (Non-Participants): 6.636
3rd place with an average vote of 6.7419		
Views: 77 (67 during voting) Votes: 31		

图2 www.dpchallenge.com上的“挑战”主题

Fig.2 The "Challenge" theme on www.dpchallenge.com

4.2 AADB数据集

KONG等人^[14]在2016年建立了AADB数据集,包含大约10,000张图像,每张图像有两类标签:第一类是美感评分,评分范围为1—5分;第二类是美感属性,研究者借鉴传统摄影原理,设计和选择三分法则、运动模糊、平衡元素等八个美感属性对图像进行描述。

AADB数据集的图像来源于网站www.flickr.com,研究者从该网站下载大量图像并进行人工清洗,去除非照片图像,最终得到10,000张照片图像,然后五位评分者在AMT众包平台对其依次进行评估。考虑到不同评分者评估的主观差异,并为了确保在对大量图片进行评估的过程中不会出现错误注释,评分者的个体评估信息被记录下来,通过概率论和数理统计的知识计算出个体评分间的显著性差异,从而实现去除错误数据的效果。

4.3 AROD数据集

SCHWARZ等人^[15]于2017年建立了AROD数据集,其图像来源和AADB数据集图像来源一致,都来自网站www.flickr.com。但是与AADB数据集构建方法不同,SCHWARZ等人通过建立一种美感模型从而直接利用该网站图像的在线

描述信息,比如图像喜好人数和查看次数,直接推导计算出美感分数。该方法能够大批量获取图像数据和美感分数,因此AROD数据集一共包含了大约38万张照片图像,规模远远大于AADB数据集。AROD数据集示例如图3所示。

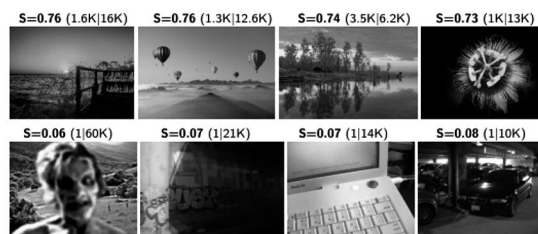


图3 AROD数据集示例^[15](S为注释分数)

Fig.3 Sample images in the AROD dataset^[15]

4.4 对比分析

如前文所述,AVA数据集中的图像来源于www.dpchallenge.com,该网站能够获取大量有美感评分的照片图像,但是该网站采用“挑战”机制,用户大多偏向上经过二次处理的美化后的照片,因此该数据集的整体图像美感质量分布并不均衡,多以二次处理后的专业照片图像为主,高美感照片图像远远多于低美感照片图像,因此无法真实地代表人类照片图像的分布情况。相较于AVA数据集,AADB数据集的照片图像更加均衡,它同时包含了普通摄影照片和专业摄影照片,且二者比例接近。并且,AVA数据集中照片图像的美感属性标签是研究者通过对每类“挑战”进行注释,从而间接对该“挑战”下的全部照片图像打上同样的二元标签,而AADB数据集则是评分者依次对每个照片图像进行评估,相比之下,AADB数据集在美感属性标签方面的注释更加准确。但是AADB数据集的总体评分人数过少,相较于AVA每张照片图像平均有210名评分者评分,AADB数据集每张照片图像仅有五名评分者进行打分,因此在评分人数方面有所不足,没有广泛的代表性。另外,AVA数据集和AADB数据集的注释都来自人工标注,而AROD数据集获取注释的方法是直接利用在线的信息得到美感分数,不需要在人工评分上花费大量精力,因此在数据集建立方面,AROD数据集提供了一种全新的思路。

与图像分类数据集相比,图像美感评估数据集的建立更加困难,因为图像分类问题是一个模拟人类客观感知的过程,而图像美感评估是模拟人类主观感知的过程。在建立照片图像美感评估数据集的时候,照片图像数量、照片图像质量的均衡程度、评分者数量和审美水平、错误标签的去除,甚至建立数据集的工作量都是需要研究者考虑的。

5 结论(Conclusion)

基于深度学习的照片图像美感质量评估作为计算美学中非常重要的研究内容,因其能够在海量照片图像中检索并推送高美感质量照片图像,在当今移动互联网时代有着广泛的应用。本文从基于深度学习的照片图像美感质量评估基本思路出发,对三种照片图像美感评估类别进行了详细的归纳。同时,按照时间顺序梳理了该领域主要的三个照片图像美感

质量评估数据集,并对其优缺点进行了分析。

数据集在照片图像美感评估中起着至关重要的作用,但由于人类评估的主观性,建立数据集的难度较大。尽管一些研究者使用迁移学习的方法对数据集的重要性进行弱化,如DONG等人^[16]提出使用AlexNet^[17]模型中倒数第二层的输出作为美感评估的特征,然后训练支持向量机对照片图像进行二元分类;LI等人^[18]使用包含两次微调的嵌入式学习,进一步利用图像分类模型提高照片图像美感评估的准确率,但是迁移学习在该领域的研究仍未达到饱和。并且,利用深度学习技术构建的神经网络学习到的特征缺乏可解释性,如何进一步深入探究照片图像美感评估工作的学习过程,是相关研究者需要思考解决的问题。

参考文献(References)

- [1] 朱立元.美感大辞典[M].上海:上海辞书出版社,2010:1.
- [2] HOENIG F. Defining computational aesthetics[C]// NEUMANN L, SBERT M, GOOCH B, et al. Computational Aesthetics in Graphics, Visualization and Imaging. Girona, Spain: The Eurographics Association, 2005:13–18.
- [3] LU X, LIN Z, JIN H L, et al. Rating pictorial aesthetics using deep learning[J]. IEEE Transactions on Multimedia, 2015, 17(11):1.
- [4] TIAN X L. Aesthetic evaluation of images based on deep and shallow feature fusion network models[C]// LIU N S, WANG Y S, FU Y G, et al. EITCE 2020: 2020 4th International Conference on Electronic Information Technology and Computer Engineering. New York, USA: ACM, 2020: 814–818.
- [5] MA S, LIU J, CHEN C W. A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment[C]// MORTENSEN E, YUAN J S, TUREK M, et al. IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles, USA: IEEE Computer Society, 2017:4535–4544.
- [6] MAI L, JIN H L, LIU F. Composition-preserving deep photo aesthetics assessment[C]// MORTENSEN E, YUAN J S, TUREK M, et al. IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles, USA: IEEE Computer Society, 2016:497–506.
- [7] DATTA R, LI J, WANG J Z. Algorithmic inferencing of aesthetics and emotion in natural images: An exposition[C]// ALZAHIR S, LABEAU F, MOCK K, et al. IEEE International Conference on Image Processing. Los Angeles, USA: IEEE Computer Society, 2008:105–108.
- [8] KAO Y, WANG C, HUANG K Q. Visual aesthetic quality assessment with a regression model[C]// ALZAHIR S, LABEAU F, MOCK K, et al. IEEE International Conference on Image Processing. Los Angeles, USA: IEEE Computer Society, 2015:1583–1587.
- [9] HUANG W Z, YANG P P, HUANG K Q. From classification to regression: Model transfer for visual aesthetic quality assessment[C]// KIMURA A, ALIREZA A, DENGEL A, et al. 2017 4th IAPR Asian Conference on Pattern Recognition. Los Angeles, USA: IEEE Computer Society, 2017:304–309.
- [10] TALEBI H, MILANFAR P. NIMA: Neural image assessment[J]. IEEE Transactions on Image Processing, 2018, 27(8):3998–4011.
- [11] CUI C R, LIU H H, LIAN T, et al. Distribution-oriented aesthetics assessment with semantic-aware hybrid network[J]. IEEE Transactions on Multimedia, 2019, 21(5):1209–1220.
- [12] WANG L, WANG X, YAMASAK T, et al. Aspect-ratio-preserving multi-patch image aesthetics score prediction[C]// DAVIS L, TORR P, ZHU S C, et al. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Los Angeles, USA: IEEE Computer Society, 2019:1833–1842.
- [13] MURRAY N, MARCHESOTTI L, PERRONNIN F. AVA: A large-scale database for aesthetic visual analysis[C]// MORTENSEN E, YUAN J S, TUREK M, et al. IEEE Conference on Computer Vision and Pattern Recognition. Los Angeles, USA: IEEE Computer Society, 2012:2408–2415.
- [14] KONG S, SHEN X H, LIN Z, et al. Photo aesthetics ranking network with attributes and content adaptation[C]// LEIBE B, MATAS J, SEBE N, et al. European Conference on Computer Vision. Berlin, Germany: Springer, 2016:662–679.
- [15] SCHWARZ K, WIESCHOLLEK P, LENSCH H. Will people like your image? Learning the aesthetic space[C]// KOVASHKA A, BOWYER K, MORTENSEN E, et al. IEEE Winter Conference on Applications of Computer Vision. Los Angeles, USA: IEEE Computer Society, 2018:2048–2057.
- [16] DONG Z, SHEN X, LI H Q, et al. Photo quality assessment with DCNN that understands image well[C]// HE X J, LUO S H, TAO D C, et al. International Conference on Multimedia Modeling. Berlin, Germany: Springer, 2015: 524–535.
- [17] KRIZHEVSKY A, SUTSKEYER I, HINTON G. ImageNet classification with deep convolutional neural networks[J]. Advances in Neural Information Processing Systems, 2012, 25:1097–1105.
- [18] LI Y X, PU Y Y, XU D, et al. Image aesthetic quality evaluation using convolution neural network embedded learning[J]. Optoelectronics Letters, 2017, 13(6):471–475.

作者简介:

李文轩(2001–),男,本科生.研究领域:计算美学,计算机图形学.