

Analyzing Emotional Semantic of Abstract Art Using Low-Level Image Features

(He Zhang, Eimontas Augilius, Timo Honkela, Jorma Laaksonen, Hannes Gamper, and Henok Alene)
(Advances in Intelligent Data Analysis X, IDA 2011)

摘要

在这项工作中，作者基于传统的二分类框架下的低级图像特征，研究人们在观看抽象艺术图像时所引起的情绪。本实验使用的是抽象艺术图像，而非真实的摄影图像，因为后者包含的上下文信息会以高度个性化的方式影响情绪的评估。例如一幅猫或山的图像引起的情绪反应是消极还是积极，是较为主观的，本研究旨在研究影响情绪的共性因素。通过实验实证地证明，通过机器学习算法，结合各种低级的图像描述符如颜色、形状和纹理，可以合理地预测抽象艺术图像所引发的情感。

关键词：情感语义；心理物理评估；图像描述符；二分类；机器学习

1 引言

情感是人类特有的一种基于主观认知经验在面对客观事物时产生的行为反应，在心理学中，情感是指与情绪相关的内部身体表征的精神对应物。人可通过面部表情、姿态语言和其它生理参数表达情感，如惊恐时通常伴随着瞳孔放大、面部肌肉紧绷、声调变高、冒汗和心跳加快等生理反应。探究情感的心理学实验揭示了这些共性反应，给识别情感提供了理论依据。情感计算的目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来建立和谐人机环境，并使计算机具有更高的、更全面的智能。情感计算的对象可分为内在情感信号和外在情感刺激。内在情感信号基于人的生理反应，有人脸表情、动作姿态、语音以及生理信号等主要研究思路是揭示此类情感信号与情感之间的对应关系。外在情感刺激包括文本、图像和视频等，主要研究思路是预测在某一种外在情感刺激下受试者会产生何种情绪。本研究围绕外在情感刺激中的图像部分，即视觉内容情感分析（visual emotion analysis），预测和度量图像和视频为刺激物使人产生的情感反馈，分析图像视频的哪些刺激会诱发对应的情感。视觉情感分析旨在识别由情感引起的视觉内容。近年来，随着社交网络用户的日益增多，社交网络的研究受到越来越多的关注更倾向于通过视觉内容表达自己的观点和情感。VEA 有许多实际应用，例如意见挖掘、商业智能、娱乐助手、个性化情绪预测。

模型	引用	类型	情感状态/维度
Ekman	[62]	CES	开心, 悲伤, 愤怒, 厌恶, 害怕, 惊喜
Mikels	[29]	CES	愉快, 愤怒, 尊敬, 满足, 厌恶, 激动, 害怕, 悲伤
Plutchik	[63]	CES	($\times 3$ 程度) 愤怒, 期待, 厌恶, 愉快, 悲伤, 惊喜, 害怕, 信任
Parrott	[64]	CES	包含一级、二级和三级情感类别的树状层次分组
Sentiment		CES	积极, 消极, (和中性)
VA(D)	[65]	DES	V-A(-D)
ATW	[66]	DES	活跃-温度-重量

图 1: 常用的情感模型

2 相关工作

在情感识别领域, 有两类常见的研究思路, 一是基于生理信号, 二是基于视觉刺激内容。基于生理信号即在试验中收集受试者的脑电图、心电图、心率变异性等生理信号和受试者当时的情绪类别来研究。基于视觉刺激内容即使用手工特征或深度特征来识别某个刺激物对应的情感, 如一副色调较暗的图像往往给人消极的情绪。两者思路的研究侧重点不同, 由于本课题是第二种, 因此下述研究现状仅分析基于特征的情感识别。

2.1 基于手工特征

在深度学习爆发年(通常认为是 2012 年)之前, 手工特征以绝对的优势占据了情感识别领域。常见的识别过程是设计的手工特征(公认具有代表性的算子有 SFIT 等)并对图像进行特征提取, 然后放入机器学习模型, 主要使用支持向量机分类。王伟凝^[1]基于心理学和颜色理论, 设计了亮度-冷系-暖系、饱和度-冷系-暖系-对比度和对比度-锐度三种图像特征, 以此挖掘图像的情感信息, 达到 87.3% 的准确率, 并建立了基于图像情感的检索系统。Zhang^[2]等使用了十一种关于纹理、形状、颜色和边缘的特征对图像进行情感识别, 探究了视觉形状对于图像情感的影响, 取得了平均 67% 的识别准确率, 证明了形状特征对于情感预测的有效性。G.Patterson^[3]等设计了一个大规模属性数据集(SUN), 包含材料、表面特征等 102 种不同类别的属性, 通过大量实验来研究场景属性和场景类别之间的相互作用, 并建立了一种场景属性特征, 将识别范围扩大到目标所在的场景。D. Borth^[4]提出了称为 Sentibank 的大规模视觉情感本体论, 包括 1200 个概念, 每一个概念表示一个形容词-名词对, 包含丰富且清晰的语义信息, 为拟合“情感鸿沟”提供了创新性的解决思路。

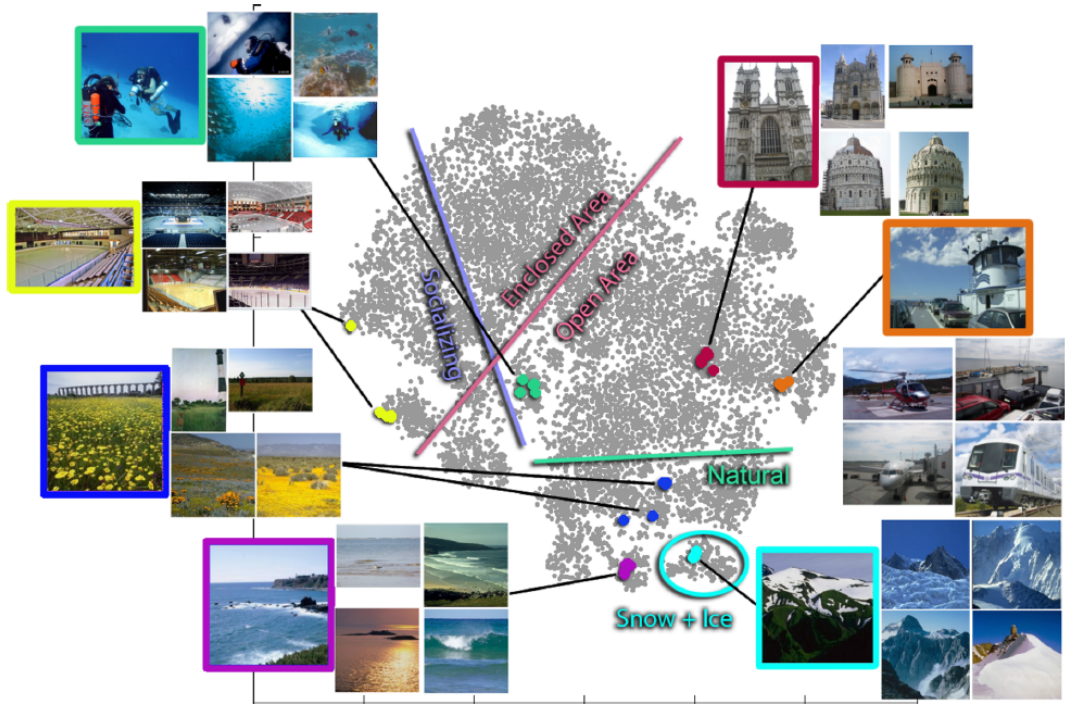


图 2: 数据集 SUN 的可视化

2.2 基于深度特征

2012 年之后, 随着卷积神经网络的快速发展和多媒体信息如视频、音频和图像等爆炸性激增, 基于学习的特征在情感图像内容分析领域展现出明显的优势。相比手工特征时代, 卷积神经网络带来的变革性影响有, 一是自动提取特征取代了设计手工特征并提取的步骤, 二是加深层网络极大提供了识别的准确率, 三是能实现更为复杂的功能, 如跨上下文多模态识别等。Zhao^[5]开发了一个极性一致的深度注意力网络 (PDANet), 将注意力整合到带有情感极性约束的 CNN 中, 并设计了一种基于弱监督情绪极性来引导注意力产生的极性一致性回归损失, 在目前细粒度视觉情感回归方面的最优方法。Yang^[6]提出了一个端到端的弱监督深度架构 (WSCNet), 将检测分支和分类分支集成统一的深度框架内, 并以端到端的方式优化网络, 经过实验表明此方法为最先进的视觉情感分析方法。Yang^[7]根据视觉情感是由物体之间的相互作用以及图像中物体与场景之间的相互作用引起这一心理学研究成果, 提出了一种场景-对象相关视觉情感推理网络 (SOLVER) 来预测图像带来的情感视觉刺激, 实验证实均优于已有的方法。

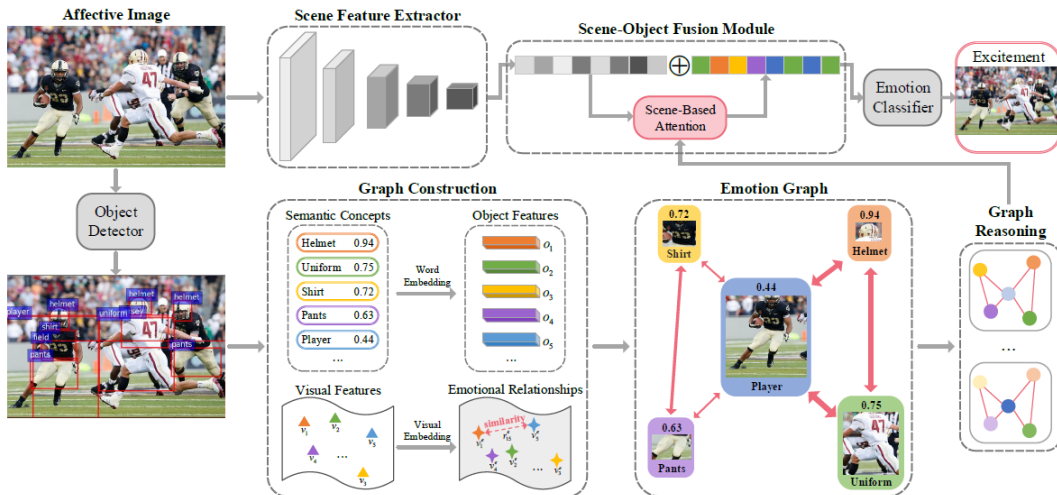


图 3: SOLVER 的网络架构

3 本文方法

3.1 数据收集

作者通过谷歌图像搜索收集了 100 幅不同尺寸和质量的抽象艺术绘画图像。这些抽象艺术作品的创作者来源各异，图像大小也在一个范围之内 185×275 和 1000×1000 。保持从互联网上选取的图像样本不变，不进行图像降采样、裁剪等数据预处理。由于原数据集数据量较小，且缺乏标签，本研究根据原实验数据集的特点，使用了 Artphoto^[8]数据集代替。ArtPhoto 数据集由 806 幅艺术图像及情感标签组成，有八类图像，分别是 Amusement、Anger、Awe、Contanement、Disgust、Excitement、Fear 和 Sad。

3.2 特征提取模块

为了描述视觉艺术画作，作者从原始图像和几种复合图像变换中提取了一大组图像特征，这些特征早先在生物图像分类和人脸识别，以及画家和艺术流派的识别中被发现非常有效^[9]。图中列出了 11 组特征。除了原始像素外，图像特征还从图像的几种变换和变换的变换中提取^[10]。这些变换是快速傅里叶变换（FFT）、小波二维分解、切比雪夫变换、颜色变换和边缘变换等。11 种特征分别是 First four moments、Haralick features、Multisacle histograms、Tamura features、Radon transform features、Zernike features、Chebyshev statistic features、Chebyshev-Fourier features、Edge statistics features、Object statistics 和 Gabor filters 等。

Group of Features	Type	Dimension
First four moments	Image Statistics	48
Haralick features	Texture	28
Multiscale histograms	Texture	24
Tamura features	Texture	6
Radon transform features	Texture	12
Chebyshev statistic features	Polynomial	400
Chebyshev-Fourier features	Polynomial	32
Zernike features	Shape & Edge	72
Edge statistics features	Shape & Edge	28
Object statistics	Shape & Edge	34
Gabor filters	Shape & Edge	7

图 4: 手工特征

3.3 实验实施

作者使用 SVM^[11]作为分类器对图像进行分类，并且默认参数，以确保实验结果的可重复性。在将图像数据集分成训练集和测试集之后，利用训练集学习 SVM 分类器。然后对于测试集中的每一幅图像，都会自动预测出一个相应的类标签或情感类别。原实验中作者对四类标签 “Exciting”、“Boring”、“Relaxing”、“Irritating” 相互两两组合，形成六种组合，并以二分类的方式通过 SVM 进行分类。

4 复现细节

4.1 特征函数的还原

由于原文提供的数据集的数据量太少，且缺乏标签，我们使用了 Artphoto 数据集来代替。里面一共有 8 个种类，按照作者的思路对它们进行两两组合共 15 种组合进行实验。复现实验的第 1 步是将作者使用的 11 种特征函数还原出来，由于没有提供源代码，只能通过查询相关论文资料来还原它，具体的做法是根据其数学含义以及数学公式来编程模拟推导。这 11 种特征由于个别缺乏明确的定义或者数学公式，无法还原，因此在本实验中一共还原出 8 种。以下是还原的八种特征的概述。

4.1.1 First four moments^[12]

包含期望值（mean）、方差（variance）、偏度（skew）、峰度（kurtosis）等概念，期望值和方差是统计学中的概念，分别表示随机变量的平均值和方差。偏度和峰度是数据分布的统计特征，偏度表示概率分布曲线的不对称程度。当分布曲线向右偏（正偏）时，数据的大部分值集中在分布的左侧，当分布曲线向左偏（负偏）时，数据的大部分值集中在分布的右侧。偏度可以通过计算样本或总体的三阶中心矩来得到。如果偏度值为零，则表示分布是对称的。峰度指的是概率分布曲线在其峰值处的峰顶尖度，即分布曲线在峰值处的陡峭程度。它反映了分布相对于标准正态分布的峰度大小，用于描述分布曲线的尖峰或扁平程度。

4.1.2 Haralick features^[13]

包含对比度（contrast）、能量（energy）、同质度（homogeneity）、相关性（correlation）、非相似性（dissimilarity）、第二角力矩（ASM）、熵（entropy）等概念。对比度、能量、同质度是图像处理中的概念，分别表示图像中不同灰度值的强度差异、图像总能量和图像内部相似程度。相关性、非相似性、第二角力矩和熵是图像特征描述中的术语，分别表示图像内部灰度值的相关性、图像不同区域的差异程度、图像内部灰度值的分布状况和图像内部随机性。

4.1.3 Tamura features^[14]

包含有粗糙度（coarseness）、对比度（contrast）、方向度（directionality）、线性度（linearity）、规整度（roughness）等。上述均是图像边缘检测和形状描述的术语，粗糙度表示图像内部粗糙程度，对比度通过对像素强度分布情况的统计得到的，其大小由四个因素决定：灰度动态范围、直方图上黑白部分两极分化程度、边缘锐度和重复模式的周期。方向度表示给定纹理区域的全局特性，描述纹理如何沿着某些方向发散或者集中的程度。线性度表示某个方向以及邻近的方向角度近似相等程度，规整度表示是表示物理表面不平整程度的一种物理量。它表示物理表面离散点的程度和分布，以及物理表面的凹凸程度。

4.1.4 Radon transform^[15]

拉东变换是一种积分变换，它将二维平面上的函数 $f(x,y)$ 沿着平面上的任意直线进行线积分，等同于对 $f(x,y)$ 进行 CT 扫描。它最常用的应用是检测图像中是否含有直线成分，因为任何直线都会在变换后产生极值。

4.1.5 Zernike features^[16]

Zernike 矩是基于 Zernike 多项式的正交化函数，所利用的正交多项式集是 1 个在单位圆内的完备正交集。当计算 1 幅图像的 Zernike 矩时，以该图像的形心为原点，把像素坐标映射到单位圆内。在模式识别中，一个重要的问题是对目标的方向性变化也能进行识别。Zernike 矩是一组正交矩，具有旋转不变性的特性，即旋转目标并不改变其模值。由于 Zernike 矩可以构造任意高阶矩，所以 Zernike 矩的识别效果优于其他方法，常用来识别目标物体形状特征比较明显的图像。

4.1.6 Chebyshev statistic features^[17]

Chebyshev 统计特征是一种在数学和计算机科学中用于图像处理、计算机视觉和机器学习等领域的特征提取方法。它是基于 Chebyshev 多项式的统计特征，通过对图像中的像素值和位置进行特征提取和表示。这种特征的优势在于可以对图像的多项式分布进行建模，并且可以使用快速算法进行计算。因此，Chebyshev 统计特征通常被用于大规模图像数据的特征提取和分类任务。

4.1.7 Gabor filters^[18]

一种对称的小波函数，它们能够同时具有时域和频域的特征可以用来提取图像中的纹理特征，例如纹理的方向和周期性，以及频率特征，例如边缘和轮廓。它们也可以用来检测图像中的纹理不连续，如边缘和轮廓。

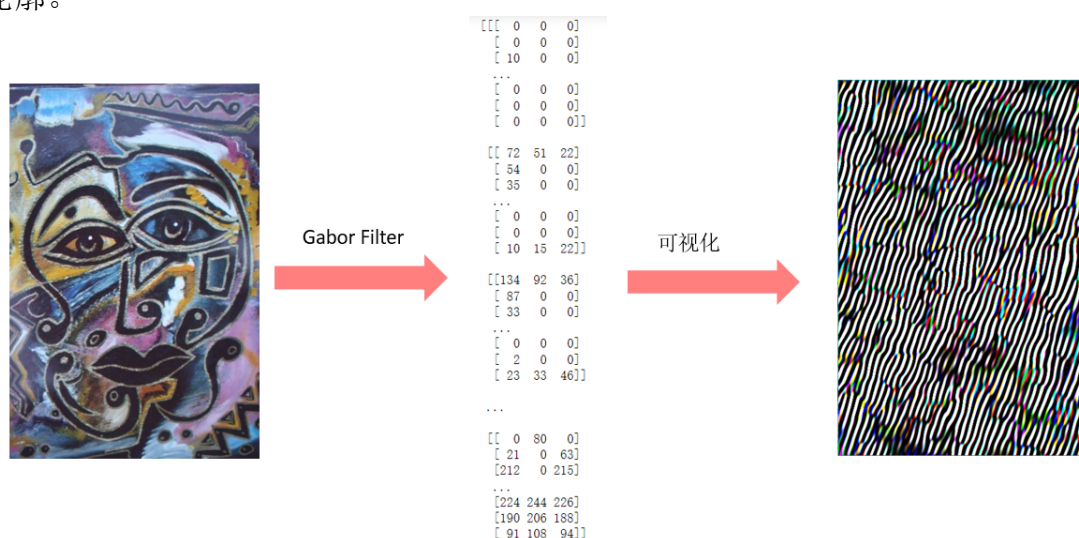


图 5: Gabor filters 的可视化

4.2 实验环境搭建

使用的编程语言：Python3.8。涉及的库：NumPy、cv2、pandas、scipy、matplotlib、skimage、mahotas、csv、pytorch、PIL、sklearn 等。代码运行环境：操作系统为 64 位 window10，处理器为 Intel(R) Core(TM)

i7-8750H CPU @ 2.20GHz 2.21 GHz, GPU 为 Intel (R) UHD Graphics630 和 NVIDIA GeForce GTX 1050Ti, IDE 为 PyCharm。

4.3 分类实验的实施

作者采用了上述定义的特征函数对所选取的所有图像进行遍历,并逐一提取其特征。提取的特征被保存至 CSV 文件中,并按照两两配对的方式将四种情绪类别组合成六种情况。使用支持向量机对这六种情况进行了二分类实验。在本研究中,我们沿袭了作者的实验思路,将八种情绪类别两两配对,产生了十六种组合。利用先前已实现的八种特征函数,我们对所选取的图像进行特征提取,并将提取的特征保存至 CSV 文件中。之后,我们将这些特征送入支持向量机进行二分类实验。

4.4 补充实验

除了上述所提到的内容,本研究还进行了一些补充实验。原实验是基于两两配对的支持向量机进行二分类,本研究补充了使用支持向量机进行多分类实验。具体而言,我们将所有提取到的特征数据及其对应的标签一次性送入支持向量机中,并观测其多分类结果。另外,我们还进行了基于深度特征的情绪分类实验。具体而言,我们使用了 Resnet101^[19]模型对该数据集进行自动特征提取,并将提取的特征送入神经网络进行多分类实验。在实验中,我们使用了预训练模型,即基于 Image net 的预训练参数。此外,我们旨在比较手工特征和深度特征在情绪识别的应用效果,以进一步探究情绪识别的最佳实践方法。除此之外我们还利用深度学习方法对情感识别领域常用的数据集如 FI^[20]、Emotion6^[21]等进行了测试。FI 是目前情感识别领域最大的数据集之一,有 23185 张图像,均为真实图像,有八个情绪类别。Emotion6 有 1982 张图像,有六个情绪种类,两者的标签均为离散型情感类别。

5 实验结果分析

5.1 原实验结果分析

原实验平均 Pre.-Real 精度达到 68%,平均 Acc.-Real 达到 67%,具有较高的应用价值。其 Relaxing-Irritating 精度最高,分别是 Pre-Real 达到 76% 和 Acc.-Real 达到 70%。最低是 Exciting-Boring,Pre-Real 为 57%,Acc.-Real 为 58%。

Case	Best	Pre.-Real	Pre.-Rand	Acc.-Real	Acc.-Rand
All 20: Exciting-Boring	27	62	33	65	48
All 20: Relaxing-Irritating	11	76	53	70	52
10 female: Exciting-Boring	3	57	49	58	50
10 female: Relaxing-Irritat.	6	70	48	69	40
10 male: Exciting-Boring	8	69	56	67	47
10 male: Relaxing-Irritat.	7	72	53	71	54
Average	10	68	49	67	49

图 6: 实验结果

5.2 本实验结果分析

本实验的二分类实验中,平均准确率为 56.3%,识别率最高的类别是 Disgust-Fear,为 62.8%,最低的类别是 Disgust-Joy,准确率为 50.3%。其中测试了不同的核函数和惩罚系数,效果最好的是使用 rbf 核函数和惩罚系数为 10。使用多分类实验中,所有类别一次送入支持向量机分类,准确率仅为 25

% 左右，效果不好。随后的基于 Resnet101 模型对 Artphoto 进行了深度特征的多分类测试，使用预训练模型后训练集精度为 94.3 %，测试集精度为 37.5 %。在 FI 数据集上训练集精度是 89.9%，验证集精度是 59.8%。Emotion6 数据集的训练集精度是 92.9%，验证集精度是 47%，均优于手工特征的支持向量机多分类结果。

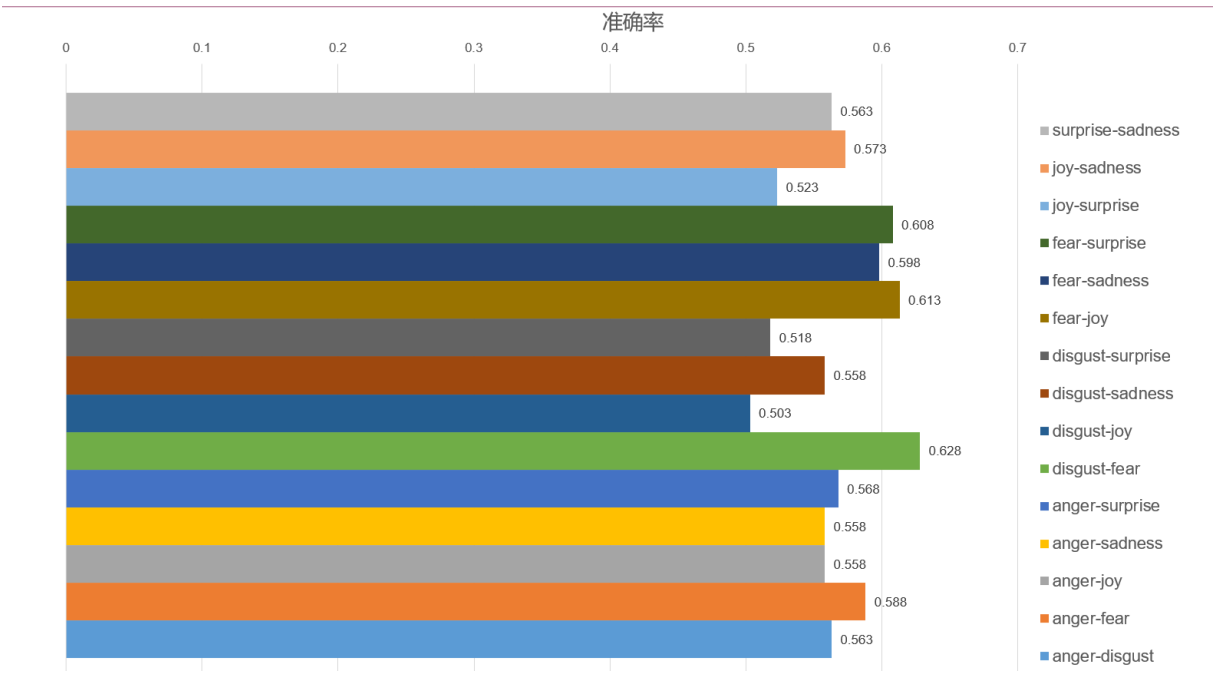


图 7: 复现实验结果

	FI	Artphoto	Emotion6
情绪种类/种	8	8	6
数量/张	23185	806	1982
轮次/epoch	200	200	200
训练集准确率	76.3%	77.1%	82.3%
验证集准确率	50%	25.9%	33.2%
Pre-trained训练集	89.9%	94.3%	92.9%
Pre-trained验证集	59.8%	37.5%	47%
Resnet-101	66.16%	69.36%	56.69%
SOTA (MDAN)	76.41%	78.12%	61.66%

图 8: 补充实验的结果

6 总结与展望

在该研究中，作者采用了数学特征函数对所选图像进行了特征提取，并使用支持向量机分类，是情感识别领域中较为经典的手工特征实验方法。在实验中，作者按照两两配对的方式组合情感类别，并使用这些向量积分类器获得了平均 67 % 的识别精度，这一方法具有不错的使用效果。本研究按照作者的思路，对所选取的情感类别进行了两两配对，并使用支持向量机进行了二分类实验。结果表明，所获得的精度为 50.3 %。随后，本研究进行了进一步的实验，使用了支持向量机和深度学习方法进行了多分类实验。在实验中，使用了预训练的 Resnet101 模型对数据集进行自动特征提取，然后将特征

数据及其对应的标签一次性送入神经网络中进行多分类实验。实验结果显示，使用支持向量机的多分类精度为 25 %，而使用深度学习方法的分类精度为 37.5 %、59.8% 和 47%。可以看出，深度学习方法^[22]在情感识别领域的应用效果优于传统的手工特征方法。本实验与原来的实验在精确度上存在一定差距。可能的原因是，一方面手工特征的还原存在不准确性，另一方面本实验未能完全还原原实验中使用的特征向量，同时使用的数据集具有一定的差异性，本实验使用的数据集是真实摄影图像，而原实验使用的数据集是创作的抽象艺术画，两个数据集所包含的色彩、边缘、纹理、形状^[23]等特征的丰富程度可能存在差异。目前在情感识别领域，主流的方法是使用基于深度学习的模型，这些模型通常具有较高的学习精度。情感识别领域正朝着越来越复杂的模型方向发展，不再仅仅是单一的图像分类问题。新的方法，如注意力机制^[24]、情感刺激区域^[25]、多网络模型等的融合^[26]，能够极大地提高学习精度。这启示我们，在当前深度学习已经被广泛应用的情况下，对于类似的分类问题，单纯依靠一个深度学习模型已经不足以取得较好的成绩，需要结合其他新的方法来提高分类效果。

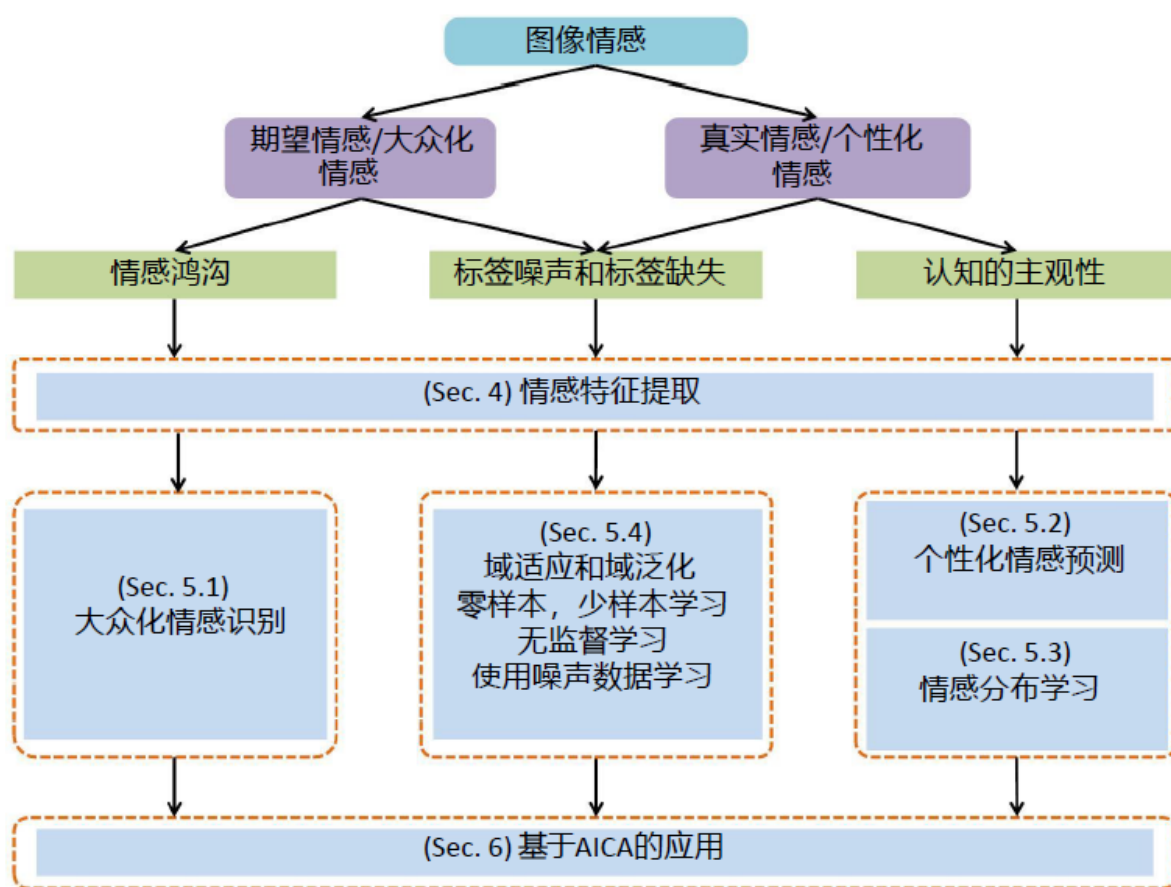


图 9: 未来的方向

参考文献

- [1] WEI-NING W, YING-LIN Y, SHENG-MING J. Image Retrieval by Emotional Semantics: A Study of Emotional Space and Feature Extraction[C]//Systems, Man and Cybernetics, 2006. SMC '06. IEEE International Conference on. 2007.
- [2] HE Z, AUGILIUS E, HONKELA T, et al. Analyzing Emotional Semantics of Abstract Art Using Low-Level Image Features[J]. Springer, Berlin, Heidelberg, 2011.
- [3] PATTERSON G, XU C, SU H, et al. The SUN Attribute Database: Beyond Categories for Deeper Scene

Understanding[J]. International Journal of Computer Vision, 2014, 108(1-2): 59-81.

- [4] BORTH D, JI R, CHEN T, et al. Large-scale visual sentiment ontology and detectors using adjective noun pairs[C]//Acm International Conference on Multimedia. 2013.
- [5] ZHAO S, JIA Z, CHEN H, et al. PDANet: Polarity-consistent Deep Attention Network for Fine-grained Visual Emotion Regression[Z]. 2019.
- [6] SHE D, YANG J, CHENG M M, et al. Wscnet: Weakly supervised coupled networks for visual sentiment classification and detection[J]. IEEE Transactions on Multimedia, 2019, 22(5): 1358-1371.
- [7] YANG J, GAO X, LI L, et al. SOLVER: Scene-object interrelated visual emotion reasoning network[J]. IEEE Transactions on Image Processing, 2021, 30: 8686-8701.
- [8] MACHAJDIK J, HANBURY A. Affective image classification using features inspired by psychology and art theory[C]//Proceedings of the 18th ACM international conference on Multimedia. 2010: 83-92.
- [9] LU X, ADAMS R B, LI J, et al. An investigation into three visual characteristics of complex scenes that evoke human emotion[C]//2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII). 2017: 440-447.
- [10] SHAMIR L, MACURA T, ORLOV N, et al. Impressionism, expressionism, surrealism: Automated recognition of painters and schools of art[J]. ACM Transactions on Applied Perception (TAP), 2010, 7(2): 1-17.
- [11] JOACHIMS T. Making large-scale SVM learning practical[R]. Technical report, 1998.
- [12] NORDHAUSEN K, OJA H, OLLILA E. Multivariate models and the first four moments[G]//Nonparametric statistics and mixture models: a festschrift in honor of Thomas P Hettmansperger. World Scientific, 2011: 267-287.
- [13] HARALICK R M, SHANMUGAM K, DINSTEN I H. Textural features for image classification[J]. IEEE Transactions on systems, man, and cybernetics, 1973(6): 610-621.
- [14] TAMURA H, MORI S, YAMAWAKI T. Textural features corresponding to visual perception[J]. IEEE Transactions on Systems, man, and cybernetics, 1978, 8(6): 460-473.
- [15] HELGASON S, HELGASON S. The radon transform: vol. 2[M]. Springer, 1980.
- [16] NOLL R J. Zernike polynomials and atmospheric turbulence[J]. JOsA, 1976, 66(3): 207-211.
- [17] MASON J C, HANDSCOMB D C. Chebyshev polynomials[M]. Chapman, 2002.
- [18] MEHROTRA R, NAMUDURI K R, RANGANATHAN N. Gabor filter-based edge detection[J]. Pattern recognition, 1992, 25(12): 1479-1494.
- [19] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

- [20] YOU Q, LUO J, JIN H, et al. Building a large scale dataset for image emotion recognition: The fine print and the benchmark[C]//Proceedings of the AAAI conference on artificial intelligence: vol. 30: 1. 2016.
- [21] PENG K C, CHEN T, SADOVNIK A, et al. A mixed bag of emotions: Model, predict, and transfer emotion distributions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 860-868.
- [22] RAO T, LI X, ZHANG H, et al. Multi-level region-based convolutional neural network for image emotion classification[J]. Neurocomputing, 2019, 333: 429-439.
- [23] RAYAR F. ImageNet MPEG-7 Visual Descriptors-Technical Report[J]. arXiv preprint arXiv:1702.00187, 2017.
- [24] ZHAO S, MA Y, GU Y, et al. An end-to-end visual-audio attention network for emotion recognition in user-generated videos[C]//Proceedings of the AAAI Conference on Artificial Intelligence: vol. 34: 01. 2020: 303-311.
- [25] YANG J, LI J, WANG X, et al. Stimuli-aware visual emotion analysis[J]. IEEE Transactions on Image Processing, 2021, 30: 7432-7445.
- [26] ZHAO S, YAO X, YANG J, et al. Affective image content analysis: Two decades review and new perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6729-6751.