

基于两层迁移卷积神经网络的抽象图像情感识别

杨子文¹, 陈蕾^{1,2,3}, 浦建宇¹

(1. 南京邮电大学计算机学院, 江苏南京 210023; 2. 江苏省无线传感网高技术研究重点实验室, 江苏南京 210023;
3. 南京航空航天大学计算机科学与技术学院, 江苏南京 210016)

摘要: 为弥合抽象图像底层视觉特征与高层情感语义间的鸿沟, 同时缓解抽象图像情感识别所固有的小样本缺陷, 将两层迁移学习策略引入传统的卷积神经网络, 提出一种基于两层迁移卷积神经网络的抽象图像情感识别模型. 该模型利用深度特征的层次性, 首先通过大规模通用图像数据集来学习提取普适的底层图像特征; 然后利用抽象图像风格分类数据集来学习提取抽象图像的专有高层语义特征; 最后采用抽象图像情感识别数据集来微调整个网络. MART 数据集上的实验结果表明, 与传统的抽象图像情感识别方法相比, 所提出的模型能够有效地提高识别精度.

关键词: 情感识别; 深度学习; 迁移学习; 卷积神经网络; 抽象图像

中图分类号: TP391 **文献标识码:** A doi: 10.3969/j.issn.0253-2778.2019.01.006

引用格式: 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.

YANG Ziwen, CHEN Lei, PU Jianyu. Recognizing emotions from abstract paintings using convolutional neural network with two-layer transfer learning scheme[J]. Journal of University of Science and Technology of China, 2019, 49(1):40-48.

Recognizing emotions from abstract paintings using convolutional neural network with two-layer transfer learning scheme

YANG Ziwen¹, CHEN Lei^{1,2,3}, PU Jianyu¹

(1. School of Computer Science, Nanjing University of Posts and Telecommunications, Nanjing 210023;
2. Jiangsu High Technology Research Key Laboratory for Wireless Sensor Networks, Nanjing 210023;
3. College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016)

Abstract: In order to bridge the gap between low-level visual features and high-level emotional semantics, and to alleviate the defects inherent in small sample dataset in abstract paintings emotions recognition datasets, a two-layer transfer learning strategy is introduced into traditional convolutional neural networks and a model for recognizing emotions from abstract paintings is proposed using convolutional neural network with a two-layer transfer learning scheme. According to the hierarchical nature of deep features, a large-scale generalized image dataset is used to learn how extract universal low-level image features. Then the relevant domain dataset is utilized to learn how extract specific high-level semantic features. Finally the abstract painting emotion recognition dataset is used to finetune the network. As shown by our extensive experimental validation on MART datasets, the proposal outperforms current methods when recognizing

收稿日期: 2018-06-14; **修回日期:** 2018-09-18

基金项目: 国家自然科学基金(61572263); 江苏省自然科学基金(BK20161516); 中国博士后科学基金(2015M581794); 江苏省博士后科研资助计划(1501023C)资助.

作者简介: 杨子文, 男, 1994年生, 硕士生. 研究方向: 机器学习和数据挖掘. E-mail: yangziwen1994@gmail.com

通讯作者: 陈蕾, 博士/副教授. E-mail: chenlei@njupt.edu.cn

emotions from abstract paintings.

Key words: emotion recognition; deep learning; transfer learning; convolutional neural network; abstract paintings

0 引言

随着文化与科技日益紧密地联系在一起,许多博物馆也试图通过多媒体的方式将馆藏的文物、绘画等艺术作品展示在互联网上^[1]. 绘画是艺术家与观者之间情感交流的桥梁. 通过计算机自动识别绘画的情感语义具有重要的意义:一方面,博物馆不仅能够有效地管理海量的数据,而且能减少领域专家的介入,从而节省人力物力;另一方面,用户能够快速检索相关绘画作品,并将相同情感的作品联系起来,方便解读绘画的含义. 与传统绘画不同的是,抽象艺术家直接用颜色、形状和纹理等视觉元素以“非具象”的方式表达情感:“艺术家试图在作品中仅表达内在真理,因此放弃对外部形式的所有考虑”^[2]. 如何弥合底层视觉元素与高层情感语义之间的鸿沟,成为抽象图像情感识别的一个研究难题.

传统的抽象图像情感识别方法大多将艺术理论和计算机视觉技术相结合,使用人工设计的特征配合统计机器学习方法来识别抽象图像唤起的情感反应. Yanulevskaya 等^[3]基于整体图像特征使用支持向量机(support vector machine, SVM)进行情感分类. Sartori 等致力于设计不同的视觉特征(如色彩组合特征^[4]、纹理特征^[5]等)结合 Group Lasso 和线性 SVM 进行抽象图像情感识别. 基于 Sartori 的工作,本文研究了鉴赏者在欣赏一幅抽象图像时所产生的情感反应是积极的还是消极的,如图 1 所示. 随着大规模高质量标记数据集(如 ImageNet^[6])的出现,深度学习技术在图像认知层识别中取得突破性成果^[7-12]. 相较于人工设计的特征,深度学习特有的多层结构,使其能够学习到逐渐从普适的底层视觉特征(如边缘、纹理等)过渡到高层语义表示(躯干、头部等)的深度特征,且层次越深,表达能力越强. 深度特征的层次性为弥合语义鸿沟、实现理解情感的认知提供了一条行之有效的途径,因此许多研究者也将深度学习用于自然图像的情感识别^[13]. 与自然图像不同的是,由于抽象图像数据集的标注需要聘请高水平专家以及设计专业的标注工具,从而导致了数据集中可用于训练的样本数量不足以训练一个深层的深度学习模型;且小样本数据集会导致深度

学习模型出现严重的过拟合问题.



图 1 抽象图像情感识别任务示例

Fig. 1 Examples of recognizing emotions from abstract paintings

解决小样本数据集下深度学习对学习任务的过拟合问题是计算机视觉领域的一个研究热点^[14]. 将从一项任务学到的知识应用于其他相关任务的迁移学习技术已被证明是有效的^[15]. 深度学习中,微调预先在 ImageNet^[6]或 Places^[16]数据集上训练的深度网络是学习特定的深度特征和缓解数据集小样本问题的常用策略,通常被认为是迁移学习技术在深度学习中的简单应用. 为弥合底层视觉特征与高层情感语义间的鸿沟以及缓解抽象图像情感识别数据集小样本问题,本文根据深度特征的层次性,提出了一个两层迁移学习策略,以渐进的方式来学习深度特征表示. 首先通过迁移大规模通用图像数据集(如 ImageNet)的权重来学习提取普适的底层图像特征;然后通过迁移抽象图像风格分类数据集(如 Painter by numbers^[17])的权重来学习提取专有的高层语义特征,最后使用抽象图像情感识别数据集来微调整个网络. 本文的主要贡献如下:

(I) 提出一种基于深度学习和迁移学习的抽象图像情感识别模型,与传统的非深度学习模型相比,本文提出的模型有效地解决了底层视觉特征与高层情感语义间的“语义鸿沟”问题;

(II) 本文提出一种两层迁移策略,根据深度特征的层次性,逐步学习大规模通用图像数据集的底层视觉特征表示和抽象图像风格分类数据集专有的高层语义特征表示,从而缓解了深度学习技术在小

样本问题上所表现出的泛化性能不足的问题;

(Ⅲ) 抽象图像情感识别数据集(MART)上的实验结果表明,相较于传统的基于人工特征的统计机器学习方法,本文提出的模型能够有效提升识别精度;同时验证了基于两层迁移策略的卷积神经网络模型比基于单层迁移策略的模型效果更好。

1 相关工作

1.1 抽象图像情感识别相关工作

先前的抽象图像情感识别工作以人工设计特征与统计机器学习方法的组合为出发点,研究了艺术作品传递给鉴赏者的情绪. Yanulevskaya 等^[3]提出了一种基于整体图像特征和 SVM 相结合的情感分类方法. Machajdik 等^[18]组合了底层视觉特征和来自心理学与艺术理论的高层语义概念使用朴素贝叶斯分类器对抽象图像进行情感分类. 文献^[19]提取了图像的 LAB 和 SIFT 特征并结合 SVM,将抽象图像根据能够引发鉴赏者的积极或消极情绪进行分类. Zhao 等^[20]提出了一种基于艺术原理的情绪特征 (principles-of-art-based emotion features, PAEF) 的图像情感分类方法,这种方法试图捕获关于图像平衡、和谐、多样和动态的信息并结合 SVM 进行分类. Sartori 等^[4]使用色彩组合特征并结合 Group Lasso 模型使用线性 SVM 进行抽象图像情感识别. 基于 Sartori 的工作, Alameda-Pineda 等^[21]提出了一种非线性矩阵补全的方法,使用色彩组合特征结合直推式多标记分类器对抽象图像进行情感识别。

1.2 融合迁移学习策略的深度学习相关工作

先前的研究者们虽然大多着力于研究如何设计有效的图像特征来进行抽象图像的情感识别,但是人工设计的视觉特征难以有效地克服低阶视觉特征和高阶语义特征之间的“语义鸿沟”问题. 随着卷积神经网络^[7]在计算机视觉领域的发展,“语义鸿沟”问题也得到了有效的缓解,但是训练卷积神经网络需要大量的数据作为基础,而现如今用于情绪语义分析领域的抽象图像数据集,以 MART 数据集^[19]为例,其仅有 500 个样本,以这种规模的数据集训练得到的卷积神经网络将会出现严重的过拟合. 迁移学习旨在将相关源域的“知识”传递给目标域^[15]. 微调预先训练过的模型^[22]是迁移学习在深度学习中的一个简单却有效的应用. 这种技术已成为学习特定的深度特征和缓解小样本问题的常用策略。

Yosinski 等^[23]通过逐层冻结使用大规模图像数据集预训练权重初始化的网络模型,证实了卷积神经网络权重的可迁移性. Zeiler 等^[24]提出一种新颖的可视化技术展示了卷积神经网络中每一层的权重和特征,揭示了靠近输入层的卷积层能够学习提取通用的底层视觉特征,而网络越靠近输出层,图像的特征表示就越抽象,高级语义特征越丰富,就越能够表现图像主题,在图像分类任务中的识别能力也就越强,从而进一步验证了深度特征的可迁移性. Oquab 等^[25]首先使用大规模已标注数据集对 AlexNet^[7]进行预训练以学习提取中层图像特征. 实验表明,尽管源域数据集和目标域数据集应用的任务不同,这种迁移方式仍然能够显著地提高识别准确率. Shin 等^[26]提出将微调卷积神经网络的方法运用于胸腹部淋巴结检测和间质性肺病检测,解决了相关领域医学图像数据集小样本的难题. Ng 等^[11]使用了一种两阶段微调的方法进行人脸表情识别,他们首先使用大规模图像分类数据集 ImageNet 对卷积神经网络进行预训练,然后使用来自不同领域面部表情识别数据集对整个网络模型的所有权重进行有监督的微调,最后使用目标领域的小规模面部表情识别数据集对模型再次进行微调. 与抽象图像情感识别不同,人脸表情识别对底层视觉特征不敏感,而对于抽象图片来说可能截然相反,因此保留 ImageNet 数据集的通用底层视觉特征是有必要的。

2 基于两层迁移卷积神经网络模型的抽象图像情感识别

当前抽象图像情感识别任务主要具有两大难点:①如何弥合底层视觉特征和高层情感语义的语义鸿沟;②如何解决情感识别领域抽象图像数据集小样本的问题. 为解决上述难点,本文使用基于深度学习和迁移学习的方法对抽象图像进行情感识别. 本节首先介绍传统的卷积神经网络结构;然后介绍提出的基于两层迁移学习策略的卷积神经网络结构。

2.1 传统的卷积神经网络模型

由 Krizhevsky 等^[7]提出的卷积神经网络 AlexNet 早在 ImageNet large scale visual recognition challenge (ILSVRC) 2012 竞赛中获得了优异的性能. 卷积神经网络也因此计算机视觉领域流行起来. 如图 2 所示,该网络以原始图像的像素作为输入,经过 5 层卷积层、3 层池化层和 3 层全连接层。

卷积层实现对输入数据以及特征图的特征提取;池化层根据不同的池化策略对特征图进行降采样;全

连接层置于卷积层后方,负责输入与输出的非线性映射。

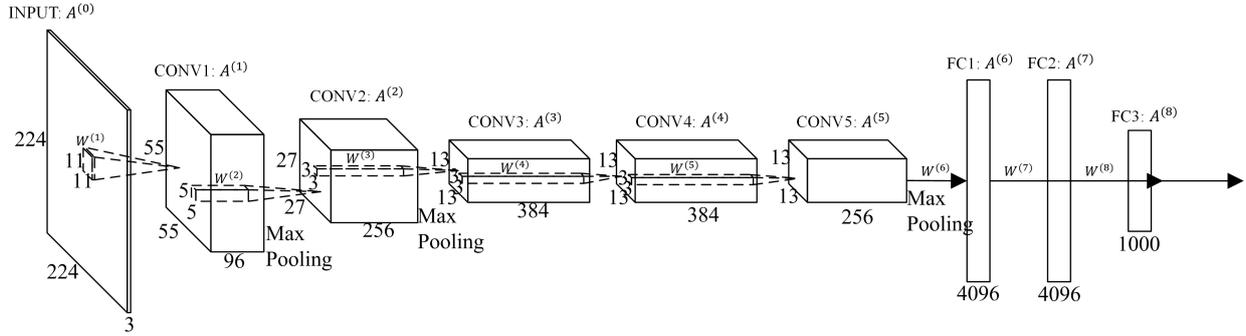


图 2 AlexNet 结构示意图

Fig. 2 A simplified illustration of the AlexNet architectures

假设 $A^{(l)}$ 为卷积神经网络中第 l 层卷积层的输出 ($A^{(0)}$ 为网络的输入图片), 则其第 i ($1 \leq i \leq M^{(l)}$) 个特征图可以表示为

$$A_i^{(l)} = \sigma \left(\sum_{j=1}^{M^{(l-1)}} A_j^{(l-1)} \otimes W_{i,j}^{(l)} + b_i^{(l)} \right) \quad (1)$$

式中, $M^{(l)}$ 为第 l 层卷积层的特征图总数, W 为卷积核权值, b 为偏置, \otimes 表示卷积操作, $\sigma(\cdot)$ 为非线性激活函数 (rectified linear unit, ReLU)^[27], ReLU 能解决梯度爆炸/梯度消失问题, 同时增强网络的表达能力^[7].

在卷积层之后可以选择性地加入池化层来对卷积层的输出进行降采样. 池化层旨在通过降低特征图的分辨率来获得具有空间不变性的特征, 且起到二次提取特征的作用^[28]. 常用的池化策略分为对卷积输出结果进行局部最大值选取的最大池化和局部平均的均值池化等, AlexNet 中使用的是最大池化层.

在卷积层和池化层将原始数据映射到隐层特征空间之后, 会连接多层全连接层将学到的特征表示映射到样本的标记空间. AlexNet 中包含了 3 层全连接层, 它们表达式为

$$A_i^{(6)} = \sigma \left(\sum_{j=1}^{n^{(5)}} W_{i,j}^{(6)} \times F_j(A^{(5)}) + b_j^{(6)} \right) \quad (3)$$

$$A_i^{(7)} = \sigma \left(\sum_{j=1}^{n^{(6)}} W_{i,j}^{(7)} \times A_j^{(6)} + b_j^{(7)} \right) \quad (4)$$

$$A_i^{(8)} = \varphi \left(\sum_{j=1}^{n^{(7)}} W_{i,j}^{(8)} \times A_j^{(7)} + b_j^{(8)} \right) \quad (5)$$

式中, $n^{(l)}$ 为第 l 层神经元个数, $F(\cdot)$ 表示将最后一层卷积层的结果展开为一维向量的平铺操作, $\varphi(\cdot)$ 为 softmax 激活函数, 用于预测输入向量属于

每一个类别的概率, 其表达式为

$$\varphi_j(X) = \frac{e^{X_j}}{\sum_k e^{X_k}} \quad (6)$$

式中, j 为输入向量 X 的第 j 个元素, k 为样本的类别总数.

CNN 的卷积操作可获得图像区域不同类型特征, 而池化、激活函数等操作可以对这些特征进行融合和抽象. 随着卷积、池化等操作的交替, 各层得到的深度特征逐渐从普适的底层视觉特征过渡到高层语义表示. 本文将使用 AlexNet 作为基础网络结构, 并结合迁移学习技术进行抽象图像情感识别.

2.2 基于两层迁移策略的卷积神经网络模型

微调预先训练过的模型是深度学习中的一个简单的迁移学习应用. 以往微调预训练模型的策略大多是使用目标任务数据集在原先预训练好的模型上继续进行训练的过程. 当源域和目标域任务相似且数据相近时, 这种方法能够有效地提高网络性能. 实际上, 大多数情况下我们难以得到与目标域数据集相似且标注质量高的大规模数据集. 受深度特征层次性的启发, 我们将传统的微调策略进一步改进为两个阶段, 如图 3 所示 (以冻结 ImageNet 前 2 层为例). 其中同一种颜色表示来自相同学习任务的结构和权重, 正方形表示输入图片, 矩形表示该层学习得到的权重, 有黑锁的矩形表示被冻结的权重, 黑色箭头表示权重迁移过程, 矩形之间的椭圆形表示网络中各个层的特征图集合, 带有网格的矩形和椭圆形分别代表卷积层的权重和特征图集合, 最后一列的圆形表示输出的概率分布.

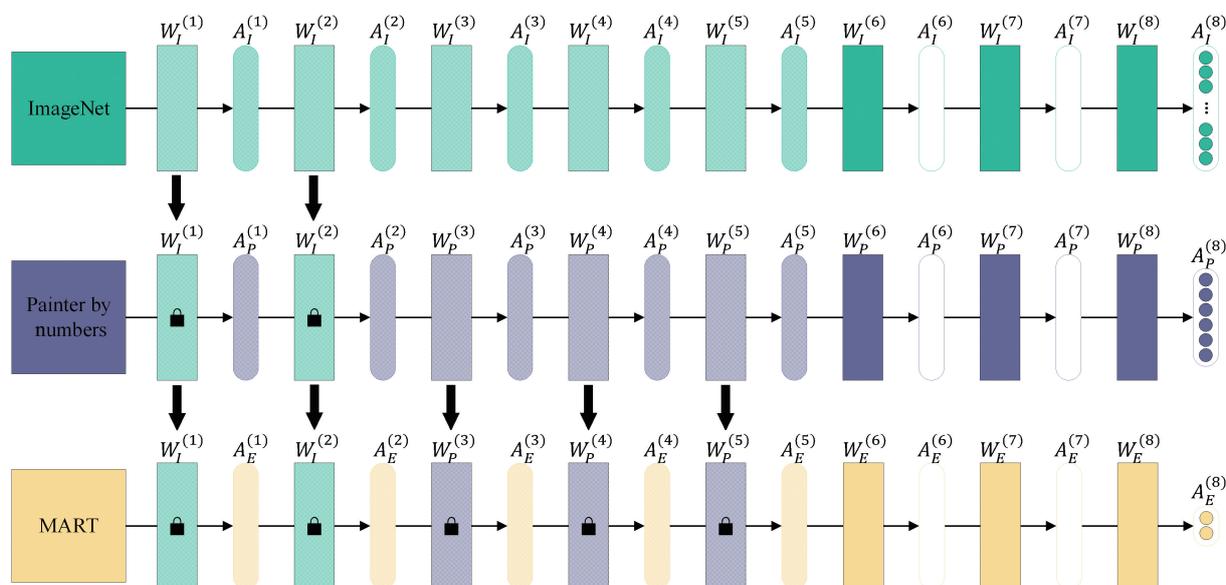


图 3 基于两层迁移卷积神经网络模型的抽象图像情感识别

Fig. 3 Overview of two-layer transfer learning strategy for the model of abstract paintings emotions recognition

第一阶段如图 3 第一行所示,我们使用大规模图像数据集 ImageNet 2012 来训练卷积神经网络. ImageNet 2012 数据集中具有超过 120 万张图片,这些图片被分为 1 000 类,其中每一类都具有 1 000 个以上的样本,其部分样例如图 4 所示,因此基于 ImageNet 数据集训练的网络模型能够得到泛化性更好的底层视觉特征.



图 4 ImageNet 数据集

Fig. 4 Examples of ImageNet dataset

第二阶段如图 3 第二行所示,根据 Yosinski 等^[23]提到的方法,首先将第一阶段中经过 ImageNet 预训练过的卷积神经网络模型的浅层卷积层(图中为前 2 层)权重,迁移到新的具有相同卷积网络结构的模型中,并冻结迁移的权重,旨在学习提取普适的底层视觉特征.与第一阶段的网络结构不同的是,第二阶段的 3 层全连接层的神经元数量

分别调整为 1 024, 1 024, 6, 以适应新的分类任务.最后使用抽象图像风格分类数据 (Painter by numbers) 对模型未冻结的部分进行微调. Painter by numbers 数据集来自于 Kaggle^[17] 竞赛数据集,它由 103 250 张艺术品图片及其对应的属性组成,包括艺术品的作者、作画时间、绘画风格等.其原始目的是为了寻找每幅作品中作者的“指纹”,即每个作者独有的绘画特征.本文只使用了此数据集关于抽象图像及其风格属性的部分,其中包括分别属于 6 种风格的 5 812 幅抽样图像作品,其部分样例如图 5 所示.第二阶段的目的是通过抽象图像风格分类数据集学习深层卷积层的权重,使得模型能够提取适用于抽象图像的高层语义特征,并且起到由底层视觉特征过渡到高层语义特征的过渡作用.

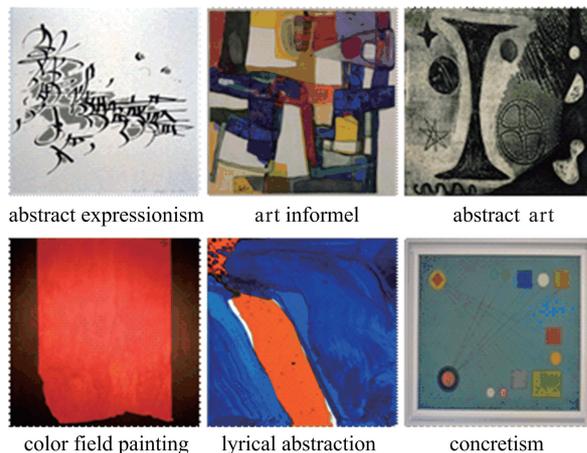


图 5 抽象图像风格分类数据集

Fig. 5 Examples of dataset for classifying abstract paintings style

最终,如图 3 最后一行所示,我们得到了基于两层迁移策略的卷积神经网络模型.该模型首先冻结第二阶段迁移后卷积神经网络的前 5 层卷积层的权重,其中前 N 层(图 3 中假设 $N = 2$)为使用 ImageNet 预训练得到的通用图像的底层视觉特征的权重,第 $N + 1$ 层到最后一层卷积层为通过抽象图像风格分类任务得到的高层语义特征的权重,随后将最后 3 层全连接层替换为神经元数量分别为 1 024,1 024,2 的全连接层以适应目标任务.

3 实验

本节将首先介绍抽象图像情感识别数据集;在 3.2 节中介绍实验设置;在 3.3 节中进行实验对比与分析.本文利用 Keras^[29]实现了如图 3 所示的基于两层迁移策略的卷积神经网络模型.

3.1 数据集

对于抽象图像情感识别任务,本文使用了被广泛用于情感分析领域的抽象图像数据集 MART^[19].MART 数据集包含了 500 张来自特兰托与罗弗莱托现代艺术博物馆(museum of modern and contemporary art of trento and rovereto, MART)的抽象艺术作品,这些作品由 78 位作家在 1913 年至 2008 年期间绘制.该数据集根据 TrueSkill 排名系统^[30]进行标注,目标是判断两幅作品“哪一幅作品更能引起您的积极情绪”,这种方式能够有效地淡化标注者的主观因素.其中每幅作品共进行 15 次比较,然后根据 TrueSkill 算法得到每幅作品的一组相对评分,最后将“更能引发观众积极情绪”8 次以上的作品分为正类,而 8 次及以下的作品分为负类.其中 MART 数据集被分为 369 个正类和 131 个负类.

3.2 实验设置

最终的抽象图像情感识别模型在卷积层使用与 Krizhevsky 等^[7]相同的超参数,而为了适应抽象图像情感识别任务,最后 3 层全连接层分别调整为 1 024,1 024,2,每个全连接层的输入都设置失活率为 0.5 的 dropout 防止模型过拟合^[14].我们使用 Adam^[31]对目标函数进行优化,并设置学习速率为 1×10^{-4} ,权重衰减为 1×10^{-6} .由于样本数量有限,所有实验参照文献[21]中的方法,使用 5 折交叉验证策略来评估模型的预测性能.具体地,整个数据集被随机分为 5 个相同大小的子集,然后依次选择每个子集作为测试数据,并将其余子集作为训练数据.

该过程独立重复 10 次,使用平均准确率(accuracy)评估最终性能.

3.3 实验结果与分析

本文首先与先前的相关工作进行了对比,其中包括 TSVM^[32](transductive support vector machine)、LMC^[33](linear matrix completion)、Lasso^[4](least absolute shrinkage and selection operator)、Group Lasso^[4]、NLMC^[21](non-linear matrix completion),这些非深度学习模型都是以文献[21]对原始图片提取颜色组合特征作为输入.同时我们还将抽象图像风格分类任务部分的权重延伸至分类器前的两层全连接层以便分析全连接部分的权重对识别准确率的影响,我们将其命名为“Proposed + FC”.由表 1 可以看出,相对于非深度学习方法,本文提出的模型在 MART 数据集上的识别精度提升了 4.8%以上.同时,直接使用随机初始化权重的 AlexNet 进行抽象图像情感识别能够获得相较于传统统计机器学习方法更高的识别准确率,但是由于小样本数据集导致的过拟合问题,其准确率远低于基于两层迁移策略的卷积神经网络.

表 1 抽象图像情感识别平均准确率(%)
Tab. 1 The average accuracy of abstract paintings emotions recognition (%)

Method	MART
TSVM	69.2
LMC	71.8
Lasso	68.2
Group Lasso	70.5
NLMC	72.8
AlexNet	73.4
Proposed+FC	76.6
Proposed	77.6

为了分析两层迁移策略的有效性,我们比较了迁移不同层数的 ImageNet 权重和 Painter by numbers 权重后模型的性能.实验结果如图 6 所示,其中 N 为迁移 ImageNet 权重的卷积层层数,即模型的前 N 层卷积层的权重为所迁移的 ImageNet 权重, N 到最后一层卷积层的权重为所迁移的 Painter by numbers 权重.故当 $N = 5$ 时表示所有卷积层的权重皆为 ImageNet 权重,当 $N = 0$ 时表示所有卷积层的权重皆为抽象图像风格分类任务的

权重.

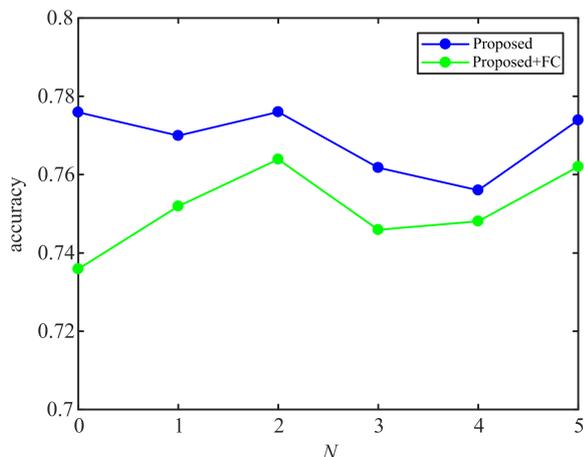


图 6 冻结前 N 层 ImageNet 权重后不同模型在 MART 数据集上的平均准确率(%)

Fig. 6 The average accuracy of different models on the MART dataset after freezing the weights of the first N convolutional layers trained by ImageNet (%)

通过观察我们可以发现:

(I) 当只对卷积层进行迁移时(即浅色曲线), 我们可以发现, 当 $N=1, 3, 4$ 时模型的准确率比 $N=0, 2, 5$ 时低, 这是因为出现了文献[23]所提到的相邻卷积层之间相互关系被破坏的情况; 当 $N=2$ 即迁移 ImageNet 前两层权重时, 通用图像的底层视觉特征能够和高层语义特征高度匹配, 此时模型的性能比 $N=0$ 和 $N=5$ 时高, 即当迁移的底层视觉特征与高层语义特征得到适当的过渡时, 基于两层迁移策略的卷积神经网络的性能优于只使用单次迁移的模型;

(II) 当将全连接层的权重也进行迁移时(即亮色曲线), 可以观察到模型的准确率出现了下滑, 这是因为不同任务的差异影响了模型的性能. 其中最明显的体现在于, 当 $N=0$ 时(即迁移 Painter by numbers 除最后一层全连接层外的所有权重)其识别准确率明显低于基于两层迁移策略的模型; 而 $N=5$ 时(即迁移 ImageNet 除最后一层全连接层外的所有权重), 相对于基于两层迁移的卷积神经网络模型, 识别准确率仅有轻微下降. 这也说明了, 不考虑源域和目标域的数据分布差异和任务差异, 单纯地使用大规模通用数据集进行微调的作用有限.

3.4 卷积核可视化

为更加直观地理解本文所提出的两层迁移卷积神经网络, 本文采用梯度上升^[34]方法为每个卷积核

生成最大化其激励的图像. 图 7 展示了卷积核可视化结果, 其每一列为每一层卷积层中随机选取的一个卷积核的可视化结果. 由图 7 可以发现, 第一层卷积层中的卷积核对颜色以及一些方向边缘信息等通用底层视觉特征进行编码, 第二层卷积层中的卷积核编码了类似于颜色与边缘信息简单组合形成的纹理特征, 而越高层的卷积核则越能响应更加抽象的图像特征, 类似于颜色、边缘、亮度等信息的复杂组合.

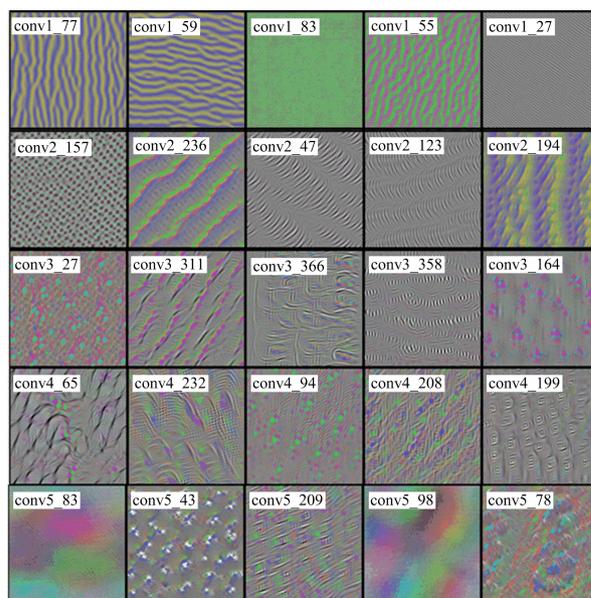


图 7 各卷积层卷积核可视化

Fig. 7 Visualization of filters in convolutional layers

4 结论

本文利用卷积神经网络深度特征的层次性, 提出了一种基于两层迁移策略的卷积神经网络模型, 有效地解决了抽象图像情感识别的语义鸿沟问题, 并缓解了小样本数据集下深度学习模型的过拟合问题. 该模型首先逐步迁移通用图像数据集的底层视觉特征和相关领域数据集的高层语义特征, 然后使用抽象图像情感数据进行微调, 最后实现了对抽象图像的情感识别. 在 MART 数据集上, 相对于传统的抽象图像情感识别方法, 本文模型的识别精度有了明显的提升, 在解决小样本数据集的任务上也优于仅仅使用一次权重迁移的方法. 考虑到艺术家们用绘画表达情感的同时, 也会使用标题和描述等信息来引导和增强鉴赏者的情感, 因此如何利用这些元数据来提高抽象图像情感识别的准确率, 将是我们未来的研究方向.

参考文献(References)

- [1] 王锐. 数字博物馆资源虚拟化与数据集成方法研究 [D]. 山东: 山东大学, 2010.
- [2] WASSILY K. Concerning the Spiritual in Art [M]. New York: Courier Corporation, 2012.
- [3] YANULEVSKAYA V, VAN GEMERT J C, ROTH K, et al. Emotional valence categorization using holistic image features [C]// Proceedings of the International Conference on Image Processing. San Diego, CA: IEEE, 2008: 101-104.
- [4] SARTORI A, CULIBRK D, YAN Y, et al. Who's afraid of itten; Using the art theory of color combination to analyze emotions in abstract paintings [C]// Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane: ACM, 2015: 311-320.
- [5] SARTORI A, ŞENYAZAR B, SALAH A A A, et al. Emotions in abstract art: Does texture matter? [C]// International Conference on Image Analysis and Processing. Genoa, Italy: Springer, 2015: 671-682.
- [6] RUSSAKOVSKY O, DENG J, SU H, et al. ImageNet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115 (3): 211-252.
- [7] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [C]// Proceedings of Advances in Neural Information Processing Systems. Lake Tahoe, USA: Curran Associates Inc, 2012: 1097-1105.
- [8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [C]// Proceedings of Advances in Neural Information Processing Systems. Montreal, Canada: Curran Associates Inc, 2015: 91-99.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 779-788.
- [10] DHALL A, RAMANA-MURTHY O V, GOECKE R, et al. Video and image based emotion recognition challenges in the wild: EmotiW 2015 [C]// Proceedings of the International Conference on Multimodal Interaction. Seattle, USA: ACM, 2015: 423-426.
- [11] NG H W, NGUYEN V D, VONIKAKIS V, et al. Deep learning for emotion recognition on small datasets using transfer learning [C]// Proceedings of the 2015 ACM on International Conference on Multimodal Interaction. Seattle, WA: ACM, 2015: 443-449.
- [12] GATYS L A, ECKER A S, BETHGE M. Image style transfer using convolutional neural networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 2414-2423.
- [13] 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 30th AAAI Conference on Artificial Intelligence. Phoenix, USA: AAAI, 2016: 308-314.
- [14] SRIVASTAVA N, HINTON G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.
- [15] PAN S J, YANG Q. A survey on transfer learning [J]. IEEE Transactions on Knowledge and Data Engineering, 2010, 22(10): 1345-1359.
- [16] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, et al. CNN features off-the-shelf: an astounding baseline for recognition [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus, USA: IEEE, 2014: 512-519.
- [17] KAGGLE. Painter by numbers dataset [OL]. [2017-10-01]. <https://www.kaggle.com/c/painter-by-numbers>.
- [18] 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. Firenze, Italy: ACM, 2010: 83-92.
- [19] SARTORI A. Affective analysis of abstract paintings using statistical analysis and art theory [C]// Proceedings of the 16th International Conference on Multimodal Interaction. Istanbul, Turkey: ACM, 2014: 384-388.
- [20] ZHAO S, GAO Y, JIANG X, et al. Exploring principles-of-art features for image emotion recognition [C]// Proceedings of the 22nd ACM International Conference on Multimedia. Orlando, USA: ACM, 2014: 47-56.
- [21] ALAMEDA-PINEDA X, RICCI E, YAN Y, et al. Recognizing emotions from abstract paintings using non-linear matrix completion [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE, 2016: 5240-5248.
- [22] 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. Columbus, USA: IEEE, 2014: 580-587.
- [23] YOSINSKI J, CLUNE J, BENGIO Y, et al. How transferable are features in deep neural networks? [C]// Proceedings of Advances in Neural Information Processing Systems. Montreal, Canada: Curran Associates Inc, 2014: 3320-3328.
- [24] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]// Proceedings of European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 818-833.
- [25] OQUAB M, BOTTOU L, LAPTEV I, et al. Learning and transferring mid-level image representations using convolutional neural networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Columbus, USA: IEEE, 2014: 1717-1724.
- [26] SHIN H C, ROTH H R, GAO M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [J]. IEEE Transactions on Medical Imaging, 2016, 35(5): 1285-1298.
- [27] NAIR V, HINTON G E. Rectified linear units improve restricted Boltzmann machines [C]// Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: Omni press. 2010: 807-814.
- [28] 周飞燕, 金林鹏, 董军. 卷积神经网络研究综述 [J]. 计算机学报, 2017, 40(6): 1229-1251.
- ZHOU F Y, JIN L P, DONG J. Review of convolutional neural network [J]. Journal of computer science, 2017, 40(6): 1229-1251.
- [29] KERAS. Keras [OL]. [2016-10-20]. <https://keras.io/>.
- [30] Microsoft. TrueSkill ranking system [OL]. [2017-11-05]. <https://www.microsoft.com/en-us/research/project/trueskill-ranking-system/>.
- [31] KINGMA D P, BA J. Adam: A method for stochastic optimization [J]. 2014, arXiv preprint arXiv: 1412.6980.
- [32] JOACHIMS T. Transductive inference for text classification using support vector machines [C]// Proceedings of International Conference on Machine Learning. Bled, France: Morgan Kaufmann Publishers, 1999: 200-209.
- [33] CHEN C H, PATEL V M, CHELLAPPA R. Matrix completion for resolving label ambiguity [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE, 2015: 4110-4118.
- [34] ERHAN D, BENGIO Y, COURVILLE A, et al. Visualizing higher-layer features of a deep network[J]. University of Montreal, 2009, 1341(3): 1-13.