第47卷 第9期 计 算 机 工 程 2021年9月
Vol.47 No.9 Computer Engineering September 2021

·图形图像处理· 文章编号：1000-3428(2021)09-0227-08 文献标志码：A 中图分类号：TP391

基于双注意力多层特征融合的视觉情感分析

蔡国永[1,2]，储阳阳[1,2]

（1.桂林电子科技大学 计算机与信息安全学院，广西 桂林 541004；
2.桂林电子科技大学 广西可信软件重点实验室，广西 桂林 541004）

摘 要：为获得更具判别性的视觉特征并提升情感分类效果，构建融合双注意力多层特征的视觉情感分析模型。通过卷积神经网络提取图像多通道的多层次特征，根据空间注意力机制对多通道的低层特征赋予空间注意力权重，利用通道注意力机制对多通道的高层特征赋予通道注意力权重，分别强化不同层次的特征表示，将强化后的高层特征和低层特征进行融合，形成用于训练情感分类器的判别性特征。在3个真实数据集Twitter I、Twitter II和EmotionROI上进行对比实验，结果表明，该模型的分类准确率分别达到79.83%、78.25%和49.34%，有效提升了社交媒体视觉情感分析的效果。

关键词：社交媒体；视觉情感分析；卷积神经网络；注意力；特征融合

开放科学（资源服务）标志码（OSID）：

中文引用格式：蔡国永，储阳阳.基于双注意力多层特征融合的视觉情感分析[J].计算机工程，2021，47(9)：227-234.

英文引用格式：CAI G Y，CHU Y Y.Visual sentiment analysis based on multi-level features fusion of dual attention[J].Computer Engineering，2021，47(9)：227-234.

Visual Sentiment Analysis Based on Multi-level Features Fusion of Dual Attention

CAI Guoyong[1,2]，CHU Yangyang[1,2]

（1.School of Computer Science and Information Security，Guilin University of Electronic Technology，Guilin，Guangxi 541004，China；
2.Guangxi Key Laboratory of Trusted Software，Guilin University of Electronic Technology，Guilin，Guangxi 541004，China）

【Abstract】In order to obtain more discriminative image features and improve the accuracy of sentiment classification，a visual sentiment analysis model based on multi-level feature fusion of dual attention is proposed.Firstly，the model extracts multi-channel and multi-level features of images with Convolution Neural Network（CNN）.Secondly，the extracted low-level features are given spatial attention weight with spatial attention mechanism，and the extracted high-level features are given channel attention weight with channel attention mechanism to saparately enhance the low and high level features.And finally，the enhaned high-level features and low-level features are fused to form discriminant features for training sentiment classifiers.Experiments on three real datasets TwitterI，TwitterII and EmotionROI showed that the classification accuracy of the method reached 79.83%，78.25% and 49.34%，which improved the effect of social media visual sentiment analysis.

【Key words】social media；visual sentiment analysis；Convolutional Neural Network（CNN）；attention；features fusion

DOI：10. 19678/j. issn. 1000-3428. 0058303

0 概述

随着移动终端技术的发展和可拍照设备的普及，越来越多的社交用户每天通过社交媒体创建和分享海量的视觉内容，这些用户生成的视觉内容的情感分析被广泛地应用于推荐、广告、舆情监控等领域。由于图像的低级视觉特征和高级情感语义之间存在难以逾越的“情感鸿沟”以及情感本身的复杂性和主观性，使得针对视觉内容的情感分析任务极具挑战性[1]。已有研究结果表明，图像情感与图像的不同层次的视觉特征相关，例如颜色等低层特征可以直接地刺激观看者从而产生情感，图像中的物体对象等高层特征同样可以传递情感。近年来，深度学习特别是卷积神经网络

基金项目：国家自然科学基金（61763007）；广西自然科学基金重点项目（2017JJD160017）。
作者简介：蔡国永（1971—），男，教授，主研方向为社交媒体数据挖掘、情感计算；储阳阳，硕士研究生。
收稿日期：2020-05-12 **修回日期：**2020-07-07 **E-mail：**381658180@qq.com

计算机工程 www.ecice06.com

(Convolutional Neural Network,CNN)在视觉识别领域展现出优异性能[2],基于深度学习的视觉情感分析方法也已经取得了较好的效果[3-4],但在克服“情感鸿沟”方面仍有提升空间。

现有的基于深度学习的视觉情感分析模型往往只利用网络的最后一层输出的语义向量作为特征进而训练分类器进行情感分类,忽略了CNN浅层所提取的有用信息。同时,由于情感的复杂性,图像之间情感差异往往由细微的局部特征所体现,因此需要在大量的局部特征中提取更具判别性的特征,现有的深度网络的特征提取能力仍需进一步增强。

针对以上问题,本文构建基于双注意力多层特征融合的网络模型DA-MLCNN。该模型通过设计一个多层全卷积网络提取多层次的图像特征,并引入空间注意力机制和通道注意力机制对不同层次的特征进行强化,最终融合多层注意力特征从而得到图像更具判别性的视觉特征以提升情感分类效果。

1 相关工作

视觉情感分析方法从特征提取的角度主要分为基于手工设计特征的方法和基于深度学习的方法。基于手工设计特征的视觉情感分析方法通过利用人工构造图像特征来进行图像的情感分类。早期的视觉情感分析研究主要通过统计学方法提取图像的颜色、纹理等低层特征[5]。文献[6]提取了图像的全局颜色直方图(Global Color Histogram,GCH)、局部颜色直方图(Local Color Histogram,LCH)等特征进行情感分析,证明了浅层视觉特征和情感之间存在着很大的相关性。受艺术原则和美学心理的启发,一些工作也围绕构图、平衡等艺术特征构建图像的情感表示[7]。随着研究的深入,语义内容等高层特征也被应用于视觉情感分析[8]。由于浅层信息难以弥合图像像素和情感语义之间的鸿沟,因此部分研究人员提出了构建情感预测的中间语义表示,如文献[9]提出了基于102个以预定义场景属性作为中层表示的图像情感分析算法Sentribute,文献[10]设计了基于1 200个形容词-名词对(ANP)的大规模视觉情感本体库,并训练情感探测器SentiBank来提取图像的中层表示。

随着社交媒体中图像数据的剧增,基于手工设计并提取特征的方法在面对大规模数据时具有很大的局限性[11]。卷积神经网络能够自动地从大规模数据中学习深度特征,因此研究人员开始将卷积神经网络应用于视觉情感分析任务。如文献[12]对在ImageNet上预训练的卷积神经网络进行微调后用于图像的情感分类,表明CNN模型优于以往基于手工特征的方法。文献[13]将CNN模型与支持向量机(SVM)相结合,在大规模的网络图像数据集上预测图像情感。文献[14]通过自适应增强算法组合多个BP神经网络弱分类器的输出,构建图像情感强分类器,提高了图像情感语义分类的效率。与改善图像全局表示的方法不同,部分研究致力于利用图像的区域信息或其他模态信息来改进情感分类准确率。如文献[15]利用CNN框架自动地发现图像中的情感区域,进而利用局部区域信息进行图像情感识别。文献[16]提出一种基于层次化深度关联融合网络同时学习图像视觉特征表示和文本语义表示,通过捕获视觉特征和文本情感特征之间的内部关联,从而更准确地实现图文融合社交媒体的情感预测。文献[17]提出一种多层次深度卷积神经网络框架,通过结合包括显著主体、颜色和局部特征在内的先验信息,从多个层次学习图像的情感表达。

注意力机制是深度学习用于模拟人类感知注意力的重要方法。为提高视觉任务的效果,视觉注意力在计算机视觉领域得到了广泛研究应用。如文献[18]引入了一种通道注意力机制,用于自适应地重新校准通道特征响应以进行图像分类。文献[19]在多层CNN结构上同时引入空间注意力和通道注意力用于图像字幕生成。已有部分研究将视觉注意力应用到图像情感分析任务中。如文献[20]首先以视觉属性探测器探测出图片的视觉属性,然后通过注意力模型关注图像中该视觉属性的响应区域进行情感识别。文献[21]将视觉注意力融入到卷积神经网络图像情感分类框架中,并利用图像的显著性映射指导空间注意力的生成以获得鲁棒的图像特征进行情感分类。

与已有研究不同,本文不仅考虑到图像情感在空间上的强弱差异以及不同通道特征之间的语义差异,通过空间注意力和通道注意力增强特征的表达,融合深度模型抽取的多层特征以充分利用不同层次特征的互补性,以期达到更好的情感分类效果。

2 方法描述

本文提出的基于双注意力多层特征融合的视觉情感分析模型DA-MLCNN整体结构如图1所示,主要包含多层CNN特征提取、双注意力机制、注意力特征融合的情感分类3个部分,分别表示如图1中(a)、(b)、(c)3个部分。

计算机工程 www.ecice06.com

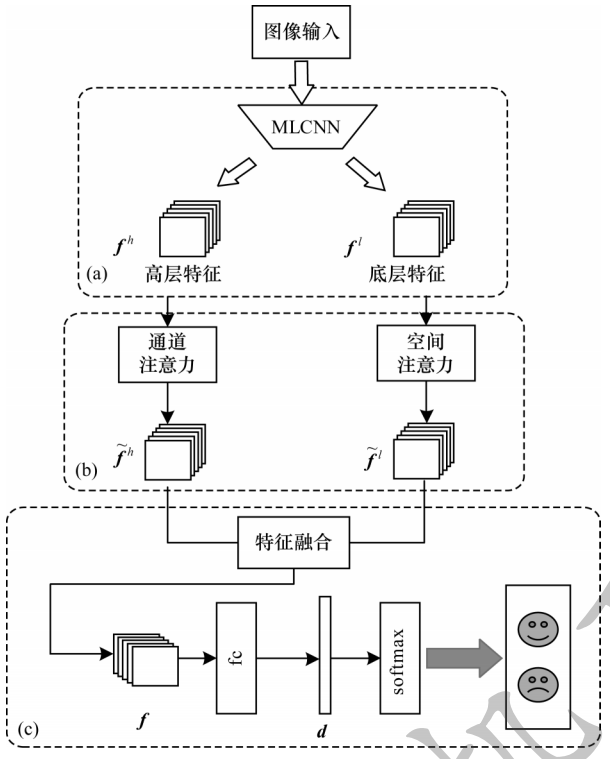


图1　DA-MLCNN模型框架

Fig.1　Framework of DA-MLCNN model

为获得更具判别性的图像特征，首先将图像输入到具有多分支结构的MLCNN模型，提取图像的不同层次的特征，即CNN浅层提取到的低层特征$\boldsymbol{f}^l$和深层提取到的高层特征$\boldsymbol{f}^h$；然后将低层特征输入空间注意力模块，将高层特征输入通道注意力模块，以关注图像中更能凸显情感的区域以及选择关注更重要的通道特征；最后融合空间注意力模块输出的结果$\tilde{\boldsymbol{f}}^l$和通道注意力模块的输出结果$\tilde{\boldsymbol{f}}^h$，形成具有判别性的特征表示用于训练情感分类器。

2.1　多层CNN特征提取

近年来通过对CNN可视化研究表明，CNN不同层提取的图像信息具有不同性质。具体而言，浅层提取到的信息更倾向于反映图像的空间信息等低层特征，而深层提取到的信息则更倾向于反映图像的语义内容等高层特征[22]。考虑到图像情感与图像的不同层次的特征相关，本文将利用不同层特征的互补性来提升情感分析性能，为此，设计一种多层全卷积网络结构MLCNN用来提取多层次的图像特征。与大部分视觉任务工作类似，本文采用近年来在计算机视觉领域应用广泛的卷积神经网络模型VGGNet-16[2]作为多层特征抽取网络MLCNN的基础。如图2所示，在VGGNet-16基础上，将网络各层输出的特征通过不同的采样方法，使其转化为相同尺寸，如将卷积层Conv1_2、Conv2_2的输出通过下采样，将Conv4_3、Conv5_3的输出通过上采样的方法，均调整为与Conv3_3相同尺寸，再将各层分别进行一次3×3卷积核的卷积操作后进行融合。其中，卷积层Conv1_2、Conv2_2的输出经过以上操作融合后的输出得到包含较多空间信息的低层特征$\boldsymbol{f}^l \in \mathbb{R}^{W\times H\times C}$，卷积层Conv3_3、Conv4_3、Conv5_3的输出经过以上操作融合后的输出得到包含较多语义信息的高层特征$\boldsymbol{f}^h \in \mathbb{R}^{W\times H\times C}$，其中，$W$和$H$表示特征图的宽和高，$C$表示特征的通道数。

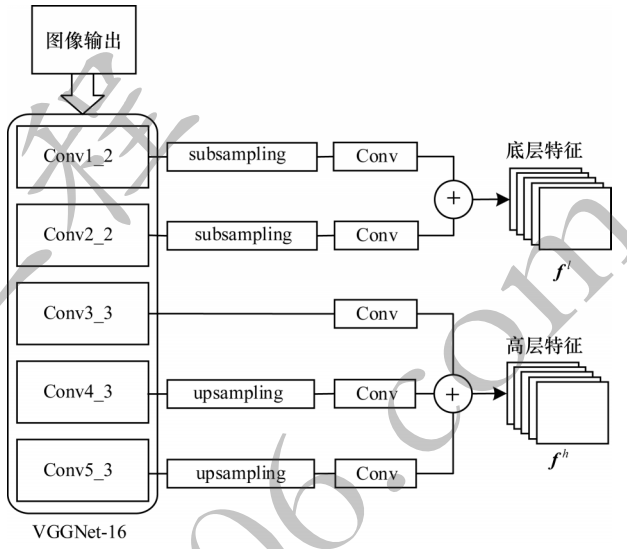


图2　多层CNN特征提取结构

Fig.2　Structure of multi-level CNN features extraction

2.2　双注意力机制

人类的视觉注意力可以突出整体特征中更重要的部分，以加强人类视觉识别的能力。注意力机制对视觉情感分析也至关重要，在视觉情感分析任务中，如果能发掘出图像中信息量最大的区域以及更重要的通道特征，将可能提升情感分类的效果。DA-MLCNN方法中的双注意力机制包括空间注意力和通道注意力2个部分。高层特征主要包含图像高层次的抽象语义，因此对高层特征使用通道注意力加强情感语义信息更丰富的特征，以此进行特征的重新标定。低层特征主要包含图像的空间信息，因此对低层特征使用空间注意力，以更加关注图像的情感区域。

2.2.1　空间注意力机制

仅利用全局图像特征向量来进行图像情感分类，结果可能不是最优的。图像中的局部区域对于图像整体的情感表达至关重要，例如图像中包含鲜花和笑脸的区域。引入空间注意力机制，通过训练方式定位出图像中凸显情感的局部区域并对其赋予更高的权重，以加强图像空间域上的特征表示。空间注意力模块结构如图3所示。

www.ecice06.com

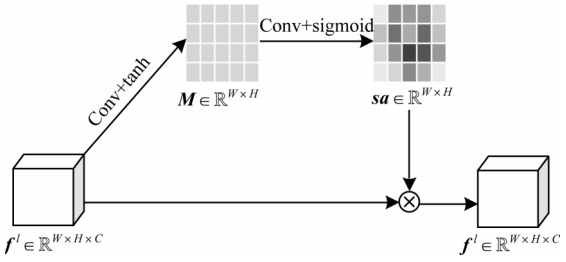


图3 空间注意力机制

Fig.3 Spatial attention mechanism

首先将低层特征$\boldsymbol{f}^l$经过1个卷积核尺寸为1×1的卷积层进行降维，然后通过tanh激活函数得到尺寸大小为$W\times H$的特征映射$\boldsymbol{M}$，再将$\boldsymbol{M}$经过卷积层和sigmoid激活函数生成空间注意力权重$\boldsymbol{sa}$，计算公式为：

$$\boldsymbol{M}=\tanh(\mathrm{Conv}_1\boldsymbol{f}^l) \tag{1}$$

$$\boldsymbol{sa}=\mathrm{sigmoid}(\mathrm{Conv}_2\boldsymbol{M}) \tag{2}$$

其中，Conv_1、Conv_2表示卷积操作。最终加权后的特征输出为：

$$\tilde{\boldsymbol{f}}^l=\boldsymbol{sa}\otimes\boldsymbol{f}^l \tag{3}$$

其中：$\otimes$表示元素相乘。

2.2.2 通道注意力机制

CNN提取的特征是多通道的，每个通道的特征表示该图片特征在相应卷积核上的分量，不同通道的特征所提取到的信息是不同的[9]，因此不同通道的特征对于情感分类的重要性是不同的。现有基于深度学习的视觉情感分析方法平等地对待每个通道特征，这在很大程度上限制了情感分类的性能。通道注意力机制关注重要的通道特征，过滤无关的通道特征，可以看做一个对特征进行重新标定的过程。本文引入了通道注意力机制，对情感有高响应的通道特征分配更大的权重以加强特征表示。通道注意力模块结构如图4所示。

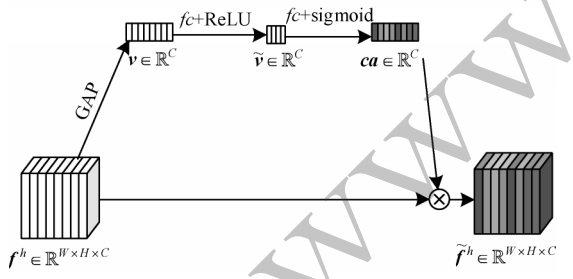


图4 通道注意力机制

Fig.4 Channel attention mechanism

对于高层特征$\boldsymbol{f}^h$，可以进一步表示为$\boldsymbol{f}^h=[\boldsymbol{f}_1^h,\boldsymbol{f}_2^h,\cdots,\boldsymbol{f}_C^h]$，其中$\boldsymbol{f}_i^h$表示第$i$个通道的特征图。首先对每个通道特征$\boldsymbol{f}_i^h$进行全局平均池化（Global Average Pooling，GAP）操作以聚合每个特征通道的信息，从而将高层特征$\boldsymbol{f}^h$压缩为C维特征向量$\boldsymbol{v}$，其第i个元素的值为：

$$v_i=\frac{1}{W\times H}\sum_{m=1}^{H}\sum_{n=1}^{W}f_i^h(m,n) \tag{4}$$

其中：$f_i^h(m,n)$表示第i个通道的特征图上位置(m,n)的值。为了生成通道注意力权重，先将$\boldsymbol{v}$通过全连接层和ReLU激活层得到$\tilde{\boldsymbol{v}}$，其维度变为C/r，r为全连接层神经元数量，再通过全连接层调整维度为C，最后通过sigmoid函数为每个通道生成注意力权值$\boldsymbol{ca}$。计算公式为：

$$\tilde{\boldsymbol{v}}=\mathrm{ReLU}(fc_1\boldsymbol{v}) \tag{5}$$

$$\boldsymbol{ca}=\mathrm{sigmoid}(fc_2\tilde{\boldsymbol{v}}) \tag{6}$$

其中：fc_1、fc_2表示相应的全连接层。最终加权后的特征输出为：

$$\tilde{\boldsymbol{f}}^h=\boldsymbol{ca}\otimes\boldsymbol{f}^h \tag{7}$$

2.3 注意力特征融合的情感分类

为了获取判别性的视觉情感特征表示，将经过注意力机制强化后的高低层次特征通过特征融合层进行融合，如图5所示。具体地，先对输入的低层特征和高层特征分别进行1×1卷积操作，得到尺寸一致的特征映射$\boldsymbol{l}$和$\boldsymbol{h}$。再对$\boldsymbol{l}$和$\boldsymbol{h}$进行拼接融合成判别性特征。令表示融合后得到$\boldsymbol{f}$的判别性特征：$\boldsymbol{f}=\boldsymbol{l}\oplus\boldsymbol{h}$，其中，$\oplus$表示特征张量的拼接操作。

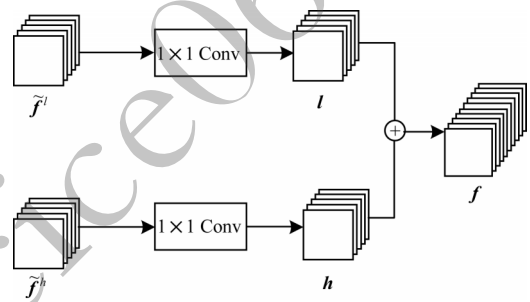


图5 多层特征融合过程

Fig.5 Process of multi-level features fusion

对于融合注意力特征后得到的判别性特征$\boldsymbol{f}$，进一步经过全连接层输出一维的语义向量$\boldsymbol{d}$。最终连接一个softmax分类层输出所有情感类别对应的概率p_i，其计算公式如下：

$$p_i=\frac{\exp(d_i)}{\sum_{i=1}\exp(d_i)},i=1,2,\cdots,m \tag{8}$$

其中：m表示情感类别。采用交叉熵损失作为整个模型的损失函数，定义如下：

$$L=-\sum_i y_i\log_a p_i \tag{9}$$

其中：y_i表示图像的真实情感类别标签，可以通过随机梯度下降（SGD）算法最小化损失函数L，梯度可通过式（10）计算：

$$\frac{\partial L}{\partial d_i}=\frac{\partial L}{\partial p_i}\frac{\partial p_i}{\partial d_i}=-\frac{y_i}{p_i}(p_i(1-p_i))=p_i-y_i \tag{10}$$

www.ecice06.com

2.4　模型训练流程

本文提出的DA-MLCNN模型的训练具体流程为：

1）通过在ImageNet数据集上进行预训练的VGGNet-16网络的权重参数初始化DA-MLCNN模型，以加快其收敛速度。

2）从所有训练样本中采样，获取训练的batch。

3）对于每个batch的样本，通过模型的多层CNN特征提取网络计算每张图像的低层特征$\boldsymbol{f}^l$和高层特征$\boldsymbol{f}^h$。

4）计算经过空间注意力机制对低层特征强化得到特征表示$\tilde{\boldsymbol{f}}^l$，计算经过通道注意力机制对高层特征强化得到特征表示$\tilde{\boldsymbol{f}}^h$。

5）融合强化后的注意力特征$\tilde{\boldsymbol{f}}^l$和$\tilde{\boldsymbol{f}}^h$，得到特征映射$\boldsymbol{f}$。

6）将特征映射$\boldsymbol{f}$经过全连接层，生成语义向量$\boldsymbol{d}$。

7）通过softmax分类器预测图像情感，并根据损失函数计算当前损失。

8）反向传播损失到每一层，并逐层更新该层对应的参数。

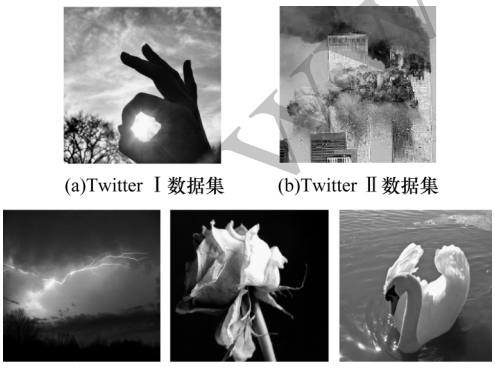
3　实验与结果分析

3.1　数据集

对本文所提出的方法在3个数据集TwitterI[4]、TwitterII[10]和EmotionROI[23]上进行了评估，以证明其有效性。实验中3个数据集均采用随机划分的形式划分为训练集80%和测试集20%。

Twitter I是从社交软件Twitter上收集的1 269张情感图片，其中标签为积极的769张，标签为消极的500张。Twitter II包含了从Twitter中收集的603张情感图片，其中积极的图片470张，消极的图片133张。EmotionROI数据集是从社交网站Flickr收集的1 980张情感图像，一共包含6类情感，即恐惧（fear）、悲伤（sadness）、愤怒（anger）、厌恶（disgust）、惊喜（surprise）、喜悦（joy）。

图6给出3个不同数据集的图片样本及其对应情感标签的示例。



(a)Twitter Ⅰ数据集　(b)Twitter Ⅱ数据集

(c)EmotionROI数据集

图6　不同数据集的示例

Fig.6　Sample of different datasets

3.2　实验设置

实验的基础网络为VGGNet-16，为了能够迁移VGGNet-16的权重参数，该部分的设置与VGGNet-16模型保持一致，所有卷积层均采用大小为3×3的卷积核，步长均设置为1，池化层采用2×2的最大池化，步长均设置为2，并在ImageNet进行预训练。输入图片为224×224的RGB彩色图像，为了防止过拟合，对每张图像样本进行5个位置的裁剪并对其进行随机水平翻转，从而对数据集进行扩充。对于每次输入的batch-size，分别设置为4、8、16、32、64、128并进行对比。采用随机梯度下降算法对网络进行优化，学习率设为0.001。模型采用Dropout策略和L2范式防止过拟合，权重衰减设为0.000 5，Dropout值设为0.5。实验开发环境为Linux-Ubuntu14.04，Python 2.7，Tensorflow 1.3.0，开发工具为PyCharm。模型的训练和测试均在Tesla P100-PCIE GPU工作站上完成。

3.3　对比方法

为证明所提出方法的有效性，本文在相同的数据集上对比了基于手工设计特征的方法和基于深度学习的方法。

1）GCH。由文献[6]提出，融合由64位二进制RGB直方图组成的全局颜色直方图特征GCH和局部颜色直方图特征LCH作为图像特征训练SVM分类器进行情感分类。

2）SentiBank。由文献[10]提出，该方法通过训练得到视觉概念检测器，可以检测视觉图像中的1 200个形容词名词对（ANP），并利用其检测结果作为特征进行情感分类。

3）DeepSentiBank。由文献[11]提出，作为SentiBank的改良版，该方法利用卷积神经网络在大规模数据集上训练的视觉情感概念分类器进行情感分类。

4）PCNN。由文献[4]提出，首先通过社交网站收集的大规模情感图像训练CNN，并采用渐进学习的策略对网络进行微调，然后通过少量人工标注的情感图像数据对模型进行诱导域转移，以提高其图像情感分类性能。

5）VGGNet-16。由文献[2]提出，首先在ImageNet数据集上进行预训练，然后利用情感图像数据集对网络进行微调。

6）COIS。由文献[24]提出，该方法通过目标检测框架对设置的5类对象进行检测，并提取图像的目标区域特征，进一步融合CNN提取的图像整体特征训练softmax分类器进行情感分类。

以上对比方法中的前2种GCH和SentiBank均为基于手工设计特征的视觉情感分析方法，后4种DeepSentiBank、PCNN、VGGNet-16和COIS为基于深度学习的视觉情感分析方法。

计算机工程 www.ecegc.com

3.4 结果分析

对于网络输入的不同batch-size设置，在TwitterI数据集下进行了对比实验。实验对比了不同batch-size值下，训练进行200个epoches所需的时间及其准确率，结果如表1所示，其中对时间进行了归一化处理。

表1 Twitter I数据集下不同batch-size值对结果的影响

Table 1 Effect of different batch-size values on results under Twitter I dataset

batch-size值	时间	准确率/%
4	7.744	64.93
8	4.344	69.35
16	2.741	74.07
32	1.653	79.83
64	1.371	77.23
128	1.000	66.98

从表1可以看出，随着batch-size的增大，处理相同数据量的速度变快，迭代相同次数所需时间减少；当batch-size为32时，达到分类准确率的最优。

表2给出了本文方法与5种对比方法在2个二分类情感图像数据集Twitter I和Twitter II上的分类结果，分类结果以准确率进行评估，准确率定义为正确分类的样本数量占总样本数数量的比例。

表2 不同方法在Twitter I、Twitter II数据集上的分类准确率

Table 2 Classification accuracy of different methods on Twitter I and Twitter II datasets %

方法	Twitter I	Twitter II
GCH	66.43	66.79
SentiBank	66.59	65.93
DeepSentiBank	71.24	70.23
PCNN	76.36	76.83
VGGNet-16	75.49	71.79
COIS	78.90	75.81
DA-MLCNN	79.83	78.25

由表2可知，本文提出的方法在TwitterI和TwitterII数据集上的分类准确率分别达到了79.83%和78.25%，准确率均高于基于手工设计特征的视觉情感分析方法GCH和Sentbank。本文提出的方法同样对比了基于深度学习的视觉情感分类模型DeepSentiBank、PCNN、VGGNet-16和COIS。由表2可知，基于深度学习的方法在图像情感分类性能上均优于基于手工设计特征的方法。而本文提出的方法在3个数据集上的分类效果相较于对比方法均有提升，对比效果最好的COIS模型在2个数据集上的分类准确率分别提升了0.93和1.42个百分点。

图7给出了本文方法与5种对比方法在多分类情感图像数据集EmotionROI上的分类结果，分类结果以准确率进行评估。

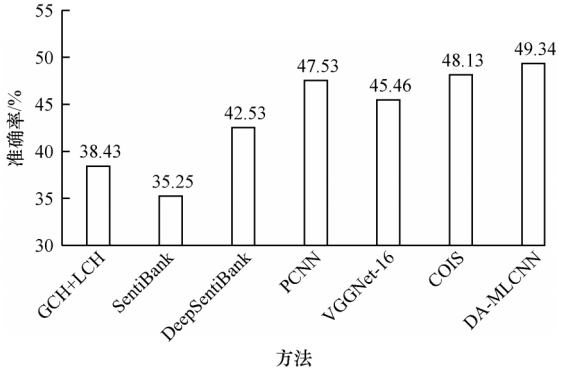


图7 不同方法在EmotionROI数据集上的分类准确率

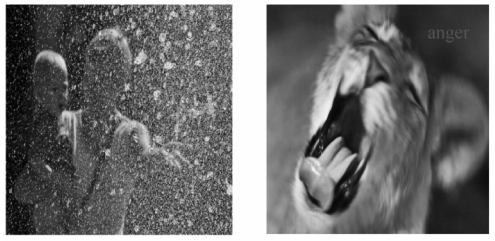
Fig.7 Classification accuracy of different methods on EmotionROI dataset

由图7可知，本文方法在多分类情感图像数据集EmotionROI上分类准确率达到了49.34%，高于基于手工设计特征的视觉情感分析方法GCH和Sentbank。对比基于深度学习的视觉情感分类模型，DA-MLCNN比DeepSentiBank和VGGNet-16的分类准确率分别提升了6.81个百分点和3.88个百分点，比PCNN和COIS模型的分类准确率提升了1.78和1.21个百分点。通过对比各类视觉情感分类方法在多分类数据集上的分类结果，可以说明本文提出的DA-MLCNN方法同样能够适应视觉情感的多分类任务。综合在二分类和多分类情感图像数据集上的分类表现，说明本文提出的方法能够学习更具判别性的视觉特征，从而提升视觉情感分析效果。

图8给出了利用本文所提出的方法正确分类的样例以及错误分类的样例。



(a)正确分类图片样例



(b)错误分类图片样例

图8 实验结果样例

Fig.8 Examples of experimental results

计算机工程 www.ecice.com

在图8中，图8(a)均被正确分类，图8(b)中2张图片分别被预测为joy类和anger类，而其真实标签分别为surprise和fear，故为错误分类。

3.5　多层特征及注意力机制的有效性分析

为进一步验证特征融合以及注意力机制在视觉情感分类任务中的作用及有效性，本文进行了消融实验，对比本文所提出方法DA-MLCNN及其不同设置的分类效果，具体说明如下：

1）VGGNet-16。本文模型DA-MLCNN所使用的基础网络。

2）MLCNN。不使用视觉注意力机制，仅通过多层CNN特征提取网络提取图像的多层次特征，并融合多层特征进行情感分类。

3）SA-MLCNN。在MLCNN基础上，提取多层次特征，仅仅对低层特征使用空间注意力，并融合多层特征进行情感分类。

4）CA-MLCNN。在MLCNN基础上，提取多层次特征，仅仅对高层特征使用通道注意力，并融合多层特征进行情感分类。

表3所示为DA-MLCNN及其设置在不同数据集上的分类准确率结果。由表3可以看出，融合多层特征后的MLCNN模型的分类准确率明显高于VGGNet-16的分类准确率，可以验证多层CNN特征融合能够充分利用不同层次特征的互补优势，提高情感分类效果。由表3可知，SA-MLCNN、CA-MLCNN、DA-MLCNN相较于MLCNN，其分类准确率分别有不同程度的提升，其中DA-MLCNN准确率提升最高，说明空间注意力和通道注意力在本文所提出的模型中能够实现更细化的特征抽取，从而加强特征表示，进而提高视觉情感分类效果。

表3　不同数据集上的分类准确率

Table 3　Classification accuracy of different datasets %

方法	Twitter Ⅰ	Twitter Ⅱ	EmotionROI
VGGNet-16	75.49	71.79	45.46
MLCNN	76.90	73.10	45.86
SA-MLCNN	77.23	75.86	48.74
CA-MLCNN	77.10	73.73	47.27
DA-MLCNN	79.83	78.25	49.34

4　结束语

现有的基于深度学习的视觉情感分析方法不能充分利用多层次的特征且模型对特征的表达能力不足。为此，本文提出一种基于视觉注意力和多层特征融合的视觉情感分析方法用于视觉情感分析。通过卷积神经网络提取图像的多层次特征，考虑到高层特征和低层特征的特点，利用空间注意力加强低层特征挖掘特征的空间信息，运用通道注意力加强高层特征关注语义内容更丰富的特征。最后融合高层特征和低层特征，形成语义增强的判别性视觉特征用于训练情感分类器。通过在3个公开的情感图像数据集上进行的实验对本文提出的DA-MLCNN模型进行评估，结果表明，本文方法能够提取更具判别性的特征表示，从而提升视觉情感分析性能。考虑到情感是复杂多样的，当前视觉情感分析的分类方法并不能反映情感的复杂性，下一步将设计更合理的网络模型提取图像特征以及研究视觉情感的细粒度分类，同时，随着社交媒体视频内容的增多，针对视频的情感分析工作也是未来的一个重要研究方向。

参考文献

[1] MACHKJDIK J，HANBURY A. Affective image classification using features inspired by psychology and art theory[C]// Proceedings of the 18th ACM International Conference on Multimedia. New York，USA：ACM Press，2010：83-92.

[2] SIMONYAN K，ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[EB/OL]. [2019-10-14]. https://arxiv. xilesou. top/pdf/1409. 1556. pdf.

[3] 高玮军，杨杰，张春霞，等. 基于AT-DPCNN模型的情感分析研究[J]. 计算机工程，2020，46(11)：53-60.
GAO W J，YANG J，ZHANG C X，et al. Sentiment analysis research based on AT-DPCNN model[J]. Computer Engineering，2020，46(11)：53-60.(in Chinese)

[4] YOU Q，LUO J，JIN H，et al. Robust image sentiment analysis using progressively trained and domain transferred deep networks[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin，Texas：AAAI Press，2015：381-388.

[5] YANULEVSKAYA V，VAN GEMERT J C，ROTH K，et al. Emotional valence categorization using holistic image features[C]//Proceedings of International Conference on Image Processing. Piscataway，USA：IEEE Press，2008：101-104.

[6] SIERSDORFER S，MINACK E，DENG F，et al. Analyzing and predicting sentiment of images on the social Web[C]// Proceedings of the 18th ACM International Conference on Multimedia. New York，USA：ACM Press，2010：715-718.

[7] 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. New York，USA：ACM Press，2014：47-56.

[8] CHEN T，YU F X，CHEN J，et al. Object-based visual sentiment concept analysis and application[C]// Proceedings of the 22nd ACM International Conference on Multimedia. New York，USA：ACM Press，2014：367-376.

[9] YUAN J，MCDONOUGH S，YOU Q，et al. Sentribute：image sentiment analysis from a mid-level perspective[C]// Proceedings of the 2nd International Workshop on Issues of Sentiment Discovery and Opinion Mining. New York，USA：ACM Press，2013：1-10.

[10] BORTH D，JI R，CHEN T，et al. Large-scale visual sentiment ontology and detectors using adjective noun pairs[C]// Proceedings of the 21st ACM International Conference on Multimedia. New York，USA：ACM Press，2013：223-232.

[11] CHEN T，BORTH D，DARRELL T，et al. Deepsentibank：

计算机工程 www.ecice06.com

visual sentiment concept classification with deep convolutional neural networks[EB/OL]. [2020-04-10]. https://arxiv. xilesou. top/pdf/1410. 8586. pdf.
[12] PENG K C, CHEN T, SAGOVNIK A, et al. A mixed bag of emotions: model, predict, and transfer emotion distributions [C]//Proceedings of IEEE CVPR'15. Piscataway, USA: IEEE Press, 2015: 860-868.
[13] YOU Q Z, LUO J B, JIN H L, et al. Building a large scale dataset for image emotion recognition: the fine print and the benchmark[C]//Proceedings of Thirtieth AAAI Conference on Artificial Intelligence. Phoenix, Arizona: AAAI Press, 2016: 308-314.
[14] 曹建芳,陈俊杰,李海芳. 基于 Adaboost-BP 神经网络的图像情感分类方法研究[J]. 山西大学学报(自然科学版), 2013, 36(3): 331-337.
CAO J F, CHEN J J, LI H F. Research on image sentiment classification based on adaboost-BP neural network[J]. Journal of Shanxi University(Natural Science Edition), 2013, 36(3): 331-337. (in Chinese)
[15] YANGJ, SHE D, SUN M, et al. Visual sentiment prediction based on automatic discovery of affective regions[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2513-2525.
[16] 蔡国永,吕光瑞,徐智. 基于层次化深度关联融合网络的社交媒体情感分类[J]. 计算机研究与发展, 2019, 56(6): 1312-1324.
CAI G Y, LV G R, XU Z. A hierarchical deep correlative fusion network for sentiment classification in social media[J]. Journal of Computer Research and Development, 2019, 56(6): 1312-1324. (in Chinese)
[17] 王伟凝,李乐敏,黄杰雄,等. 基于多层次深度卷积神经网络的图像情感分类[J]. 华南理工大学学报(自然科学版), 2019, 47(6): 39-50.
WANG W N, LI L M, HUANG J X, et al. Image emotion classification based on multi-level deep convolution neural network [J]. Journal of South China University of Technology(Natural Science Edition), 2019, 47(6): 39-50. (in Chinese)
[18] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [C]//Proceedings of IEEE CVPR'18. Piscataway, USA: IEEE Press, 2018: 7132-7141.
[19] CHEN L, ZHANG H, XIAO J, et al. SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning[C]//Proceedings of IEEE CVPR'17. Piscataway, USA: IEEE Press, 2017: 5659-5667.
[20] YOU Q, JIN H, LUO J. Visual sentiment analysis by attending on local image regions[C]//Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, USA: AAAI Press, 2017: 231-237.
[21] SONG K, YAO T, LING Q, et al. Boosting image sentiment analysis with visual attention[J]. Neurocomputing, 2018, 312: 218-228.
[22] ZEILER M D, FERGUS R. Visualizing and understanding convolutional networks [C]//Proceedings of European Conference on Computer Vision. Berlin, Germany: Springer, 2014: 818-833.
[23] PENG K C, SADOVNIK A, GALLAGHER A, et al. Where do emotions come from? predicting the emotion stimuli map[C]//Proceedings of IEEE ICIP'16. Piscataway, USA: IEEE Press, 2016: 614-618.
[24] CAI G, HE X, PAN J. Visual sentiment analysis with local object regions attention[C]//Proceedings of International Conference on Pioneering Computer Scientists, Engineers and Educators. Singapore: [s. n.], 2019: 479-489.

编辑 索书志

计算机工程 www.ecw.com